

THE ECONOMIC ANALYSIS OF LABOR MARKET DISCRIMINATION: A SURVEY

GLEN G. CAIN*

University of Wisconsin-Madison

1. Introduction

This survey of the economics of labor market discrimination is motivated by two fundamental problems associated with income and wage differences among groups classified by sex, race, ethnicity, and other characteristics. The first is the inequity of long-lasting differences in economic well-being among the groups; in particular, differences in household or family income. The second is the inequity of long-lasting differences in the average wage rates among groups of workers classified by these demographic traits, when the groups may be presumed to be either equally productive or to have equal productive capacity. The second problem also raises the question of whether a labor market that pays unequal wages to equally productive workers is inefficient.

Economic discrimination is defined in terms of income differences among families and wage differences among workers. In Section 2, I discuss these definitions and present data from the United States on the income and earnings differences of blacks, Hispanics, whites, women, and men.

Section 3 surveys theories of economic discrimination in the labor market. The theories are classified into competitive and monopolistic neoclassical models with (essentially) complete information, competitive neoclassical models with imperfect information – leading to “statistical discrimination”, and institutional theories. Only neoclassical models offer generalizable theories that can be rigorously tested, but I argue that these theories lack supporting empirical evidence.

*I am grateful to many persons, including most of the authors discussed in the chapter, for comments, criticisms, and corrections. Even though I did not always follow their advice, the chapter has been much improved because of their help. I am especially indebted to the following persons, who read and commented on the entire manuscript: Francine Blau, Betty Evanson, Ross Finnie, Arthur Goldberger, James Heckman, and Elizabeth Uhr. Research support was received from the Institute for Research on Poverty at the University of Wisconsin-Madison and from the U.S. Department of Health and Human Services. Any opinions expressed here are my own.

Empirical tests of the economic theories are selectively surveyed in Section 4. Most attention in this section is, however, given to a survey of the estimations of wage (or earnings) functions for various groups of workers as a way of measuring labor market discrimination, operationally defined as differences in predicted wages (for the groups) when the prediction "holds constant" various productivity determinants of wages.

A distinction is made between marketwide estimates of labor market discrimination and estimates that apply to an individual firm. Both methods commonly use multiple regression, but they differ primarily in the specification of exogenous predictor variables – that is, variables that may be assumed to affect wages but not to reflect the process of discrimination. The statistical models of discrimination in individual firms have become widely used in recent years as evidence in court cases or other litigation stemming from antidiscrimination laws. Although the estimates of predicted wages in both firms and markets contain much useful information, there are inherent weaknesses in the models in terms of interpreting the estimates as measures of labor market discrimination.

The chapter concludes with a discussion of the policy implications of the economic research on discrimination. Data on the changes over time in comparative earnings of women and men and of black men and white men are used as a basis for discussing the role of policies in explaining and affecting these changes.

2. The definition of economic discrimination

2.1. *Concepts*

Economic discrimination is a concept that defies precise definition. One difficulty is that the intended meaning of the term differs in several contexts in which it is used. To define economic discrimination I proceed in steps and begin with two problems that span the economist's scope of interests and expertise, from the practical to the theoretical.

(1) A practical problem, based on observed and quantified outcomes in the economy and of intense concern to the public at large, is the wide disparity in income, earnings, and wage rates among a variety of demographic groups, classified by sex, race, ethnicity, and other characteristics. The disparities are systematic, persistent, and considered by most observers to be inequitable, although the definitions and sources of the inequities are often controversial. For brevity, I will refer to the group experiencing lesser economic rewards as the "minority" group and the more favored group as the "majority" group. The fact that discrimination, in the sense of disparate outcomes and inequitable treatment, has been alleged to affect many different groups complicates its conceptual definition and makes a review of empirical work overwhelming. In this chapter I

concentrate on discrimination in the United States against women, who are not a numerical minority group, and blacks. References to discrimination against certain ethnic groups, age groups, and the handicapped will sometimes be made to elucidate certain general issues.

(2) The theoretical problem, which might be purely hypothetical except that it has been motivated by the first problem, is: *Under what conditions will essentially identical goods have different prices in competitive markets?*¹ In practice, the question refers to goods that are, on average, the same and to a price difference that is sustained rather than transitory. Economic discrimination refers to a group rather than to an individual, and it is of greater concern as it persists over time. This theoretical problem may be specified more rigorously, but let us first consider its constituent parts to see its practical implications.

Discrimination in the labor market takes labor services as the good in question and the wage rate as the price. Labor services are considered "essentially identical" if they have the same productivity in the "physical" or "material" production process; a consideration that excludes the effect of the laborer on the psychic utility of his or her coworkers or employers. In fact, psychic disutility is an essential part of a useful definition of economic discrimination that was formulated by Becker (1957, rev. 1971) and which will be discussed in Section 3. If the employers, for example, feel a disutility in hiring a minority worker solely because of the worker's demographic characteristic, which, by itself, is irrelevant to the worker's physical productivity, then employers may be said to be prejudiced. As another example, if the majority group of coworkers manifest their feelings of psychic disutility by actions which curtail the minority worker's physical productivity, this outcome will still be considered discriminatory, because the operative or causal variable is the majority group's prejudices, not the minority group's productivity. Under some but not all conditions, these tastes, which reflect the prejudices of employers and workers, will lead to discrimination, defined by wages to the minority group being below what they would receive if only their physical productivity were determinant. There is, therefore, a distinction between discrimination, which refers to behavioral outcomes, and prejudice, which refers to attitudes. My point is not that tastes are the sole source of discrimination; rather that they not be allowed to define away discrimination.

The concept of physical productivity, although it excludes the psychic component, is intended to be broad and to include such characteristics of the workers as their regularity in attendance at work, dependability, cooperation, expected future productivity with the firm, and so on. A grey area occurs when there is customer contact with the workers and when it is the customers who feel the

¹ Gustav Cassel, the renowned Swedish economist, may have been the first to state this question in the context of labor market discrimination in his analysis in 1918 of why women doing similar work to men received lower wages. See the citation to Cassel along with an interesting discussion of the history of the economists' debate on labor market discrimination in Lundahl and Wadensjö (1984, pp. 8–80).

psychic disutility. Here, the distinction between physical and psychic components of production can break down. "Service with a smile is our product", may be the company's motto. It will be argued below that although customer prejudice can lead to discriminatory outcomes, it is unlikely to be a major source of the economywide disparities in the wages and incomes between minority and majority groups.

Implicit in the foregoing two concepts of economic discrimination are two subclassifications that are defined by the unit of the analysis; namely, (a) the household (or family), which is generally the appropriate unit for examining the disparity between majority and minority groups in economic well-being, usually measured by income; and (b) the individual worker, the appropriate unit for examining disparities in wage rates or earnings. In most of this survey the worker is the unit of analysis, reflecting the fact that labor market discrimination, measured by wage disparities, has been the focus of most economic-theoretical and econometric studies. Nevertheless, attention to the family as a unit and income as an outcome is important. The family is the principal matrix for a worker's choices, and an understanding of labor market discrimination requires attention to this family context. This is most clearly evident in analyzing discrimination against women. Also, our ultimate interest in labor market discrimination lies in the question of how discrimination affects the economic well-being of people, which, as noted, is most meaningfully measured for a household or family unit.

Each of the two units of observation, worker and household, may be analyzed with two general types of statistical models. In Model (I), which may apply to the short run, the outcome variable of interest – income for households or wages for workers – is compared for the two groups, holding constant certain variables that are believed (a) to affect the outcome variable (or to be relevant to the interpretation of the outcome variable), and (b) to be exogenous to the process of discrimination under study. For example, income of households may be compared, holding constant the region of residence. If region of residence is exogenous and the cost of living varies across regions, then income is a better measure of economic well-being when region is held constant in the comparison. If region of residence is endogenous to the process of discrimination, then it is probably not a proper control variable.

Model (I), which is distinguished by the use of control variables, is more important for the second definition of discrimination – wage differences for comparable workers. The comparability of the workers is with respect to their productivity, which is operationally defined by measurable characteristics of the workers that are accepted as determining productivity in the given context. Here again we require that the productivity variables that are properly held constant are exogenous to the process of discrimination under study.

Let us specify Model (I) in a form suitable for statistical estimation. Let Y_i = the outcome of the process, such as the income, earnings, or wage for the i th

person; X_i = a vector of productivity characteristics of the i th person that are presumed exogenous in that they do not depend on Y nor on the particular form of economic discrimination under study; $Z_i = 1$ if the person is in the majority group and 0 if in the minority group; e_i = a random error term; and let A and B be coefficients representing the effects on Y of Z and X . Assuming a linear and additive model for convenience and suppressing subscripts to avoid clutter, we have

$$Y = X'B + AZ + e. \quad (\text{I})$$

Then, a regression in which we find $A > 0$ would be evidence of discrimination. The contrary case is assumed to be $A = 0$, so “reverse discrimination” ($A < 0$) is not being considered. In Model (I), the two groups designated by Z are assumed to provide “essentially identical” labor services, conditional on (holding constant) X . Equivalently, we could define market discrimination, D , as

$$D = (\hat{Y}|X, Z=1) - (\hat{Y}|X, Z=0),$$

where \hat{Y} is the predicted value of Y conditional on X , so in the above linear and additive model, $D = A$.

Now consider Model (II), in which all X characteristics are considered endogenous, and any difference in X across groups is attributed to the process of discrimination under study. Model (II) may be appropriate for the long run, although some may consider it only the limiting case in which the group averages of all X 's are equalized in the long run in a world without economic discrimination. The corresponding specification is

$$Y = CZ + u, \quad (\text{II})$$

where u is a random error and $C > 0$ is evidence of discrimination. In this case, we can define $D = \bar{Y}_{\text{maj}} - \bar{Y}_{\text{min}}$, now using unconditional means instead of conditional means, substituting the mnemonic subscripts for the Z -values, and adopting notation suitable for describing samples instead of populations. The long-run model deliberately ignores the common distinction between the occurrence of discrimination “within” versus “prior to or outside” the labor market.

The practical problem of disparities in economic well-being, usually defined in terms of differences in household incomes, is generally addressed by Model (II). The practical-and-theoretical problem of differences in wages for equally productive workers is generally examined by Model (I). However, Model (I) may be specified as close to Model (II) as desired by restricting the set of admissible X characteristics.

An interesting and unusual feature of the economic analysis of discrimination is the attention given to the roles of tastes and nonpecuniary aspects in market

transactions. The economist's treatment of tastes is, however, circumscribed. Tastes are fundamentally taken as given, and explaining their sources or how they may be changed tends to be left to the other social sciences. Instead, the economist's main objective is to determine certain behavioral outcomes that are the consequences of these tastes—specifically the disparities in employment, wages, and so on. Market outcomes become indirect measures of tastes and the focus of attention. Direct measures, such as those obtained from attitudinal surveys, which are a staple in sociology and psychology, are seldom used in economics. Despite these largely self-imposed limits of the economic analyses, the goals of predicting market outcomes and predicting the effects of policies aimed at altering these market outcomes are important and difficult.

The productivity of a given worker is also influenced by the tastes of that worker. Adherence to Model (II) implies that minority and majority groups are equal in both their productive capacity *and* their willingness to produce. Equal productive capacity refers to a common presumption of innate equality among racial and ethnic groups. Innate equality in "effective" capacity may also be assumed for women, relative to men. Thus, the biological difference in physical strength between men and women may be presumed to convey no net advantage in earnings or productive capacity to men in today's labor market. Such differences clearly lead to differential sorting into specific occupations, just as they do within a gender or racial group, but there is no necessary reason for this specialization to lead to an average wage difference across groups.

Equal willingness to produce refers to equality in tastes for market work relative to leisure when comparing racial groups and to tastes for market work relative to the combined time allocation to housework and leisure when comparing men and women. Are such tastes predetermined, or are they determined, or at least affected, by discrimination? Prior equality in tastes between men and women is often denied on the grounds that cultural and biological forces, which are presumed exogenous to the economic system (or, more narrowly, to the labor market), are the causes of a preference for market work relative to housework among men and vice versa for women. In principle, Model (I) allows any X -variable, including tastes, to be correlated with gender, because the gender effect on wages is estimated net of the X 's. However, as discussed below, the choice of X 's is often disputed.

Another conventional stance taken by economists in their study of discrimination is that the state of technology is given, which is the analogue in production to the assumption of given tastes in consumption. The issue arises whenever a distinctive trait of the minority group places it at some disadvantage because of the existing state of technology. In my view, if it would be costly to change the technology to accommodate the minority group, then there is no presumption of discrimination. The minority group in this industry or firm would simply be considered less productive. If the technology is not costly to change, then the

market, in the absence of discrimination, should already have provided the accommodating change. Thus, lowering the height of shelves could equalize the productivity of those minority groups who tend to be shorter, and new construction offers the opportunity to build ramps instead of stairways to accommodate people in wheelchairs. Perhaps some market impediment, such as government regulations, might need to be eliminated to permit the accommodation. These issues are interesting, but they will not be discussed in this chapter. Technology is assumed to be exogenous, but, like physical strength, it is not considered an important source of average productivity differences between racial groups or between men and women.

The meaning and measure of income as an index of economic well-being and of the "wage rate" are complicated issues in any practical or empirical examination of either Model (I) or (II). For example, measuring the wage as the price of labor services must deal with distinctions between current and lifetime returns to work and between pecuniary and nonpecuniary returns and, at times, with the measure and evaluation of leisure and the rewards to housework. Some specific examples may be helpful. Black men appear to receive fewer nonpecuniary benefits from their market work than do white men [Lucas (1974)]. If so, the wage advantage of white workers would be even greater if the nonpecuniary aspects of employment were monetized and included. On the other hand, black men spend less time at work than white men. Does this compensate them for their lower wages and earnings? The usual answer is "yes" if the time not at work is voluntary and perhaps considered to be leisure, but "no" if it is "involuntary unemployment". The latter may create anxiety and distress for the unemployed person and have a zero or negative value. Another example concerns household work by women, for which the rewards are, let us assume, the income shared by the family unit. Does this income compensate women for their lower market earnings? The issue, discussed more fully below, partly depends on the degree to which women's allocation to housework and market work is voluntary, or, perhaps equivalently, the degree to which women's tastes for market and housework are exogenous.

The complexities in measuring the Y -outcomes as indicators of discrimination should not be overemphasized. Sometimes one measure is believed to understate, and another to overstate, discrimination, and yet both measures may give qualitatively similar results. Evidence for this outcome is provided below. Usually, the disparity remains whether the wage or income is used, and whether the wage is measured with or without an allowance for nonpecuniary aspects of the job.

In summary, measures of economic discrimination in the labor market are the positive coefficients, A and C , in Models (I) and (II), assuming the proper measure of Y , the suitable choice of one of the two models, and, if Model I is chosen, the suitable specification and measure of X . These qualifications and the

subsequent interpretation of the coefficients and their properties all require a theoretical framework, to be discussed in Section 3.

Before presenting statistical evidence on discrimination, let us note several strengths and weaknesses of the concepts used. Their strengths include their links to market-based measurements of variables that are of intense concern to the general public as well as to the technicians who study the problem. They are robust in the face of "special cases" or individual deviations, so long as these cases and deviations are "random" with respect to the process that is modeled.

This last strength, however, may be viewed as a weakness from the perspective of various ethical or legal definitions of discrimination. When Model (I) or (II) applies to a large aggregate, such as the nationwide labor market, then a finding of no discrimination on average could be consistent with many individual cases of discrimination, so long as these were balanced by a sufficient number of cases of reverse discrimination. Lawyers and philosophers need not be put out of business by findings that A or C equal zero. [The distinctions between applying Models (I) and (II) to marketwide versus, say, individual-firm contexts will be discussed in Section 4.]

Now consider that the above economic measure of discrimination is silent about segregation. Either perfect integration or complete segregation is consistent with a finding of no discrimination. In particular, the economic definition accepts "separate but equal (wages)" as no discrimination, even though segregation may be considered noxiously discriminatory in legal and ethical senses.

Using wage differences rather than segregation indices to measure discrimination in the labor market is a corollary to my emphasis on wage discrimination rather than employment discrimination in this chapter. One justification for this emphasis, in addition to the convenience of the measurability of wages, is that when discrimination takes the form of widespread refusals to hire or promote minority workers, this should lower their relative wages. The rejected minorities must bid for jobs in less favored firms, industries, occupations, and so on. This process has been referred to as the "crowding hypothesis" [Bergmann (1974)], but my point here is that wage outcomes will reflect this reduced demand for the minority group.² A second justification is that wage discrimination can exist irrespective of the degree of integration or segregation in the market. Thus, wage differentials are a more fundamental measure of labor market discrimination than are employment differentials between majority and minority groups. None of these arguments for my use of wages as a basis for measuring and discussing labor market discrimination denies that employment and hiring statistics are appropriate in many practical contexts, including court cases involving discrimination.

²"Crowding" is an old concept. Lundahl and Wadensjö (1984, p. 73, n. 16) trace it back to John Stuart Mill, and they cite F. Y. Edgeworth and Millicent Fawcett as early twentieth-century users of the term regarding labor market discrimination against women.

2.2. Summary statistics for two concepts of economic discrimination

The first definition of economic discrimination, concerning differences in economic well-being, permits a simple measure of the differences in mean household or family income. Annual money income is assumed to be the indicator of economic well-being, and the difference will be expressed as a ratio of the minority group's income to that of the majority group.

Some comparisons of the incomes in 1981 among white, black, and Hispanic households and families are shown in Table 13.1.³ The table is detailed, and it may be helpful to note the following highlights and interpretations.

(1) Blacks and Hispanics constitute about 17 percent of the U.S. population. The total numbers of households and families by ethnic status are shown in columns 6–8 in rows 1, 3, 5, 8, and 9. Along with other smaller minority groups, such as American Indians and certain Asian immigrant groups, about 20 percent of the U.S. population may be classified into ethnic minority groups that are often believed to be victims of economic discrimination.

(2) The average income of a black household, \$14 900, is 63 percent of that of a white household, which is \$23 700. (See row 1, columns 1–3.) On a per-person basis, the ratio is only 56 percent, reflecting the larger average size of black households, as shown in row 2, columns 3, 6, and 7. As discussed below, the ratio of black-to-white income has been fairly steady in recent years but has risen over a longer period of time.

(3) The ratios of black-to-white and Hispanic-to-white incomes tend to be around 0.6 or 0.7. The average income per member of a black family headed by a woman is, however, only 32 percent of the average income per member of a white married-couple family.⁴ (Using column 2, row 6, and column 1, row 4, we obtain: $2.8/8.8 = 0.32$.) This is a large difference.

(4) Poverty status for families in 1981 was officially defined to be an annual income of \$9300 or less for a family of size four and of \$7300 or less for a family of size three. Thus, a substantial proportion of black and Hispanic families headed by women are poor, whereas only a small proportion of black and Hispanic married-couple families are poor. For most minority-group families, therefore, discrimination regarding family income in the United States is not so

³ The term "white" will be used to refer to non-Hispanic whites. "Hispanic" refers to persons of Spanish origin, who may be members of any race. Persons whose origins are Mexican, Puerto Rican, Cuban, or who are from other Central or South American countries constitute most of the Hispanic group in the United States. A "household" consists of all persons who live together in a housing unit and includes one-person households. "Families" are defined as two or more persons related by blood, marriage, or adoption, and residing together.

⁴ The term "female household head" refers to a household or family in which the primary earner is usually an adult woman without a husband present. The terms "householder" and "female householder," which are currently being used in the official statistics of the U.S. government, are defined in terms of the person in the household in whose name the dwelling unit is owned or rented. Statistics for households (or families) with a female householder are nearly the same as those that would apply to the older designation, female-headed households (or families).

Table 13.1
Mean annual incomes and income ratios of white, black, and Hispanic households and families,
United States, 1981.

Demographic unit	Mean annual income (\$000's) and B/W and H/W ratios ^a					Number of units (in millions); average number of persons per unit in parentheses		
	<i>W</i> (1)	<i>B</i> (2)	B/W (3)	<i>H</i> (4)	H/W (5)	<i>W</i> (6)	<i>B</i> (7)	<i>H</i> (8)
1. Households ^b	\$23.7	\$14.9	0.63	\$18.4	0.77	72.8	9.0	4.0
2. (per member) ^c	8.9	5.0	0.56	5.3	0.59	(2.67)	(2.99)	(3.49)
3. Married-couple families ^d	28.7	21.9	0.76	22.1	0.77	43.3	3.2	2.3
4. (per member)	8.8	5.8	0.66	5.4	0.62	(3.27)	(3.79)	(4.07)
5. Female-headed families ^e	15.3	9.8	0.61	10.8	0.70	6.6	2.6	0.7
6. (per member)	5.4	2.8	0.52	3.1	0.58	(2.84)	(3.50)	(3.47)
7. Female-headed families as proportion of all families ^f						0.12	0.41	0.23
Families with primary earner working "full time": ^g								
8. Married-couple families ^h	30.5	25.9	0.85	22.3	0.73	27.8	1.9	1.5
9. Female-headed families ^h	18.0	13.4	0.74	15.9	0.88	2.5	0.8	0.2

Source: U.S. Bureau of the Census, Current Population Reports, Series P-60, No. 137, *Money Income of Households, Families, and Persons in the United States: 1981* (Washington, D.C.: U.S. GPO, 1983), Tables 4, 13, and 19.

^aIncomes are rounded to the nearest hundred, but the ratios are based on unrounded incomes. For example, the original mean household incomes for whites and blacks in the first row are \$23 742 and \$14 856.

^bHouseholds consist of all persons who live together in a housing unit and include one-person households.

^cMean annual income per member is money household income divided by the average size of the household. For example, for white households: \$23 742/2.67 = \$8892, which, rounded and expressed in thousands of dollars, is 8.9.

^dThe Census Bureau defines a family as two or more persons related by blood, marriage, or adoption, and residing together. In this table, married-couple families do not include the relatively small number of families in which the wife is listed as the owner of the housing unit. When the wife is listed as the owner, the family is classified under "female householder". The term "householder" has replaced the term "headship" in government tables.

^eDoes not include the relatively small number of female-headed families with a husband present.

^f"All families" includes the relatively small number of female-headed families with a husband present.

^g"Full time" refers to year-round, full time, defined as working 50-52 weeks for 35 or more hours per week in 1981.

^hMedian incomes are listed instead of mean incomes, which are not reported.

much a problem of poverty, at least as officially defined, as it is of inequality – their incomes are low relative to the incomes of the white majority group.

(5) One reason black and Hispanic incomes are lower is that the fraction of families headed by a woman is larger among these minority groups. If both headship status and the presence of a full-time worker are held constant, the income ratios rise to around 0.8. (See rows 7-9, columns 3 and 5.) Marital instability and slack labor markets thus appear to be important sources of

income inequality across ethnic groups in the United States. Whether marital and employment statuses should be held constant in assessing discrimination depends on the particular purposes and issues in one's analysis. As noted above, one issue is whether marital and employment statuses are affected by discrimination.

(6) This type of table is more difficult to construct for other minority groups of interest, but consider the reported incomes for the following three groups that faced discrimination in the United States in the past:

(a) persons of Italian ancestry – the largest group of immigrants to the United States in the twentieth century;

(b) persons who state their religion as Jewish, whose ancestors had immigrated predominantly from Eastern Europe; and

(c) persons of Japanese ancestry – the largest group of immigrants from Asia.

Several researchers have concluded that the average family incomes of each of these groups was, in 1970, *higher* than the average in the United States for all other white families.⁵

What adjustments to the available statistics for money income are required to measure relative economic well-being more completely? A satisfactory answer to this question would involve the resolution of philosophical and measurement problems that are beyond my capacity, but most of the issues that lend themselves to quantification or informed judgments are listed in Table 13.2. In the table the sources of inequality and the accompanying adjustments are separated into those pertaining to income receipts and those pertaining to expenditures. In measuring income receipts attention is given to (a) the measures of income from a household's assets (or wealth components); (b) the demographic unit of analysis; (c) allowances for government taxes and subsidies; and (d) allowance for survey biases.

There is not the space to discuss each of these adjustments, but two conjectures may be suggested. First, the money measures in Table 13.1 probably underestimate the true degree of inequality between blacks and whites, and, by extension, between majority and minority ethnic groups generally. Seven of the 10 required adjustments serve to widen the gap. Second, even descriptive statistics about "income differences" in discrimination studies are complicated.

Table 13.1 shows a static picture of income differences, and it is essential in an analysis of discrimination to describe how these differences have changed over time. The time-series data are, unfortunately, incomplete in several respects.

⁵It is more difficult to define and collect information on groups according to their ancestry and religion than it is for gender and racial classifications, so the statements in the text are more qualified. The problems of mixed or unknown ancestry, changes in one's religion, response refusals and errors, and so on appear serious, and the data on income, earnings, and wage rates have not been collected for ancestry and religion classifications as thoroughly as they have for the gender and racial groups. The sources for the research findings in the three ethnic groups referred to in the text are Greeley (1976, p. 52) and Sowell (1981, pp. 5, 126–127) for Italian-Americans, Chiswick (1983) for Jews, and Sowell (1981, pp. 5, 177–178) for Japanese-Americans.

Table 13.2
Sources of inequality in economic well-being, illustrated with a comparison of
black and white families in the United States.

Source	Judgment as to whether accounting for the source would widen or narrow the conventional black-white income gap. (No adjustments needed, N.A., implies that the conventional ratio already allows for the source.)
<i>Income receipts</i>	
Asset ownership	
Property (income-earning)	N.A.
Property (non income- earning: car, owner- occupied house, etc.)	Widens gap (blacks have less wealth in these types of durable goods)
Human capital (wage earnings)	N.A.
Human capital (fringe benefits and nonpecuniary aspects of work)	Widens gap ^a
Defined for "household" as unit	
Adjust for family or household size	Widens gap (unless the comparison is already "per member") ^b
Adjust for multiple earners to allow for "leisure" consumption	Narrows gap (whites have 1.65 earners per family; blacks, 1.47) ^c
Allowance for government taxes, transfers, and survey bias	
Taxes	Narrows gap (reflecting the moderate degree of progressivity in the tax system)
Money transfer payments	N.A.
Nonmonetary transfer payments to nonaged persons (primarily Food Stamps, public housing, Medicaid)	Narrows gap (about 25 percent of black and 8 percent of white families receive these forms of noncash transfers) ^d
Nonmonetary transfer payments to aged persons (medical care subsidies and various tax advantages for the aged)	Widens gap ^e
Nonmonetary public benefits (parks, police service, etc.)	Widens gap ^f
Nonreported income	?
<i>Expenditures</i>	
Discriminatory pricing – housing, capital markets, consumer credit, etc.	Widens gap ^f
Expenditures on "regrettables" – items that do not directly produce utility, such as health maintenance, transportation to work, "waiting times"	Widens gap ^{f,g}

^aFringe benefits are generally large for jobs with higher wages and salaries. For evidence that blacks have, on average, jobs with less prestige and less pleasant working conditions, see Robert E. B. Lucas, "The Distribution of Job Characteristics," *Review of Economics and Statistics*, 56 (November 1974): 530–540.

^bSee Table 13.1.

^cSource: Table 29 in source cited in Table 13.1.

^dSource: U.S. Bureau of the Census, *Current Population Reports*, Series P-60, No. 136, *Characteristics of Households and Persons Receiving Selected Noncash Benefits, 1981* (Washington, D.C.: U.S. GPO, 1983), p. 3.

^eMedical care subsidies are derived primarily from the Social Security system, and white persons benefit disproportionately for two reasons: (1) eligibility and payments tend to be positively related to earnings during preretirement years; (2) whites live longer. The tax advantages of the aged are generally greater for higher-income persons among the aged.

^fA personal judgment.

^gFor a definition and application of the concept of "regrettable" expenditures, see William N. Nordhaus and James Tobin, *Is Growth Obsolete?* (National Bureau of Economic Research 50th Anniversary Colloquium, Columbia University Press, New York, 1972).

Income statistics prior to 1940 are scanty. The Census Bureau's time series of annual family income begins in 1947, and separate income statistics for blacks begin in 1967 and for Hispanics in 1972.

The income ratios are relatively stable year by year (not shown), but the change over decades is notable. To summarize the trends, several 10-year averages of the annual ratios of minority-to-majority incomes for the period since 1947 are shown in Table 13.3. The ratio of nonwhite-to-white family income rose from 0.37 in 1939, when most blacks lived in the low-income Southern region and on farms, up to 0.6 or more in the middle 1960s, when the ratio more or less stabilized. Since then it has been held down by the increasing proportion of black families headed by women, and, probably, by the relatively high unemployment levels from 1975 on.⁶ Whatever the reason, progress regarding the first type of economic discrimination, family income differences, has been painfully slow.

Table 13.4 shows the earnings of workers instead of the incomes of families. To the extent that earnings measure the economic well-being of workers, the table shows economic discrimination according to the definition of discrimination that was based on disparities in well-being. According to the definition that was based on wage rate differences among comparable workers, Table 13.4 would provide a measure only if we considered the worker groups – three ethnic groups and two gender groups – to be equally productive.

In Table 13.4 ratios ranging from 0.5 to 0.7 characterize most of the comparisons between minority men and white men and between women and men within each ethnic group. However, minority women earn around 90 percent of the earnings of white women. The earnings ratios of women to men and of black men to white men are smaller for “all workers” than for “year-round, full-time workers” (hereafter, “full-time”), because women and black men are less likely to work full time. (The proportion of full-time workers to all workers is shown in parentheses in the first three columns of the last two rows. More young workers and higher unemployment among these minority groups are two sources of these lower proportions.)

Clearly, the earnings ratios for full-time workers are closer to the ratios of hourly wage rates, because the all-worker variation in hours worked in the definition of earnings – hours worked times the average wage per hour – is nearly

⁶Family income depends importantly on the number of earners per family, and this number has increased among white families relative to black families in the last 20 years. The main reason is that the percentage of all families headed by women rose from 21 percent in 1960 to 42 percent in 1980 among black families and by 8 percent to 14 percent among white families [U.S. Bureau of the Census (1983c, p. 54)]. Families headed by women tend to have fewer earners than married-couple or male-headed families. The change in work rates among wives, who are the largest and most important category of secondary earners in families, did not much affect the racial difference in earners per family. The rise in labor force participation rates of wives with husbands present was similarly rapid for both color groups from 1960 to 1981: from 30 percent to 50 percent for white wives and from 41 percent to 60 percent for black wives [U.S. Department of Labor (1982, p. 714)].

Table 13.3
Median family income ratios: Black-and-other races/white; black/white;
and Hispanic/white; annual averages for five periods, 1939-1982.

Year or period ^a	Black-and-other races/white ^b	Black/white ^c	Hispanic/white ^d
1939	0.37	-	-
1947-1956	0.54	-	-
1957-1966	0.54	-	-
1967-1976	0.63	0.61	0.69 ^d
1977-1982	0.62	0.57	0.68

Sources: U.S. Bureau of the Census, Current Population Reports, Series P-60, Nos. 43, 137, and 140, published in years 1964, 1983, and 1983 respectively. The full citation of No. 137, which gives the family income figures for 1947-1981 is: U.S. Bureau of the Census, Current Population Reports, Series P-60, No. 137, *Money Income of Households, Families, and Persons in the United States, 1981* (Washington, D.C.: U.S. GPO, 1983), p. 39.

^aThe years 1947-1982 are divided into four periods, and the average of the annual ratios are reported for each period. The first year for the continuous time series of annual incomes (see sources) is 1947.

^bThe category black-and-other nonwhite races is more than 90% black for most of the period and is the only category continuously available for the earlier years. Except for the recent decade or so, the trends in the ratios for nonwhites and for blacks appeared very similar, based on the scattered evidence available. In recent years, however, the proportion of blacks among the nonwhite races has declined. Also, the proportion of black families headed by women has risen most sharply during the last 10 years or so, and this has tended to make the family income statistics for blacks diverge from those of other nonwhite races.

^cThe first year in which blacks are reported separately is 1967.

^dFamily incomes of persons of Hispanic origin were first reported in the annual series in 1972; therefore, the period for the Hispanic/white ratio is 1972-1976.

eliminated. Among working women, minority women are more likely to be full-time workers, so the ratios of minority women's earnings to white women's earnings are higher for the all-worker group.

The time series of earnings ratios for full-time workers, which is shown in Table 13.5, is useful because among the available measures it comes closest to providing a comprehensive comparison for minority and majority workers of the trends in the relative price (wage) of labor services. For this interpretation, one must assume that the full-time workers remain about the same fraction of the total population of workers, or that deviations represent (a) voluntary shifts to part-time work, and (b) no systematic selection regarding workers' productivity traits, in the changing distribution of part- and full-time workers. A change in age composition could change the distribution, and, ideally, one would want to hold constant an exogenous trait like age when constructing the time series. Assuming that any group differences in these types of compositional shifts are

Table 13.4

Mean earnings, earnings ratios, and numbers of all workers and of year-round, full-time workers for men and women; whites, blacks, and Hispanics, United States, 1981.

	Numbers of all workers in millions; and year-round, full-time workers as a ratio of all workers in parentheses ^a			Mean annual earnings (\$000's) ^b			(Black/white) and (Hispanic/white) earnings ratios, by gender		(Women/men) earnings ratios, by ethnicity		
	W (1)	B (2)	H (3)	W (4)	B (5)	H (6)	B/W (7)	H/W (8)	W (9)	B (10)	H (11)
All workers:									0.48	0.69	0.59
Men	58.2	5.7	3.6	\$17.5	\$11.6	\$12.5	0.67	0.72			
Women	45.7	5.6	2.5	8.3	8.0	7.5	0.97	0.90			
Year-round, full-time workers:									0.58	0.76	0.70
Men	(0.65)	(0.58)	(0.61)	22.8	15.7	16.5	0.69	0.72			
Women	(0.44)	(0.49)	(0.45)	13.3	12.0	11.5	0.90	0.87			

Source: Table 55 in source cited in Table 13.1.

^aA year-round, full-time worker is one who works (or is paid for) 50–52 weeks and 35 or more hours per week.

^bEarnings are rounded to nearest hundred, but the ratios are based on unrounded earnings. For example, the earnings for white and black men in the first row are \$17453 and \$11629, respectively. The use of median earnings, which are about 8 percent lower, would not much change the comparisons.

minor.⁷ Table 13.5 shows gains over time in earnings ratios for black women relative to black men (column 2), black men relative to white men (column 4), and black women relative to white women (column 5). The earnings ratio of white women to white men (column 1) has been remarkably stable at around 0.6 over this 43-year span. The ratios for Hispanics (columns 3 and 6–7) are for too brief a period to measure a trend.

Further analysis of these trends will be presented later, but the following points seem evident.

(1) The ratios for the most recent period, 1975–1982, generally remain so far short of unity that “slow progress” is a fair and regrettable assessment. The exception is the remarkable rise to near-equality for black and white women, despite the fact that their earnings ratio in 1939 was the lowest one shown in the table. This rise is partly explained by the huge exodus of black women from domestic service, one of the lowest-paid occupations, and the migration of blacks

⁷The only check on these questions of compositional shift that is easily ascertained is that of the age composition. A time series of five observations from the decennial censuses from 1940 to 1980 of the percentages of the population and of the labor force that is young (age 14–24), middle-aged (25–64), and old (65 and over) show similar trends for the race and gender groups. Thus, the age factor is unlikely to be an important source of variation in the earnings-ratio trends in Table 13.5.

Table 13.5
Median-earnings ratios for year-round, full-time workers, gender and ethnicity
comparisons, annual averages for four periods, 1939–1982.

Year or period ^a	Women/men earnings ratio by ethnicity			Black/white earnings ratio		Hispanic/white earnings ratio	
	White (1)	Black (2)	Hispanic (3)	Men (4)	Women (5)	Men (6)	Women (7)
1939 ^b	0.61	0.51	—	0.45	0.38	—	—
1955–1966 ^b	0.61	0.61	—	0.62	0.65	—	—
1967–1974 ^c	0.58	0.70	—	0.68	0.83	—	—
1975–1982 ^d	0.59	0.76	0.70	0.73	0.94	0.72	0.86

Sources: Various years for the P-60 Series of the Current Population Reports. See Table 13.4 for full citation.

^aThe years 1955–1982 are divided into three periods, and the average of the annual ratios are reported for each period. The first year for the continuous time series of earnings for year-round, full-time workers is 1955, but the 1940 census provides this figure for 1939.

^bRatios are for wage and salary earnings (excludes self-employed workers) for whites and nonwhites, who are defined as blacks and other nonwhite races in later Census publications.

^cRatios are for all earnings (includes self-employed workers and self-employment income) for whites and blacks. The first year for which blacks are reported separately is 1967. The black/white ratios are, on average, about 0.01 lower than the nonwhite/white ratios for men, and about 0.02 lower for women. The trends in both ratios, black/white and nonwhite/white, are virtually identical.

^dSame as note c; also, 1975 is the first year in which earnings are reported separately for Hispanic workers.

generally from the low-income rural sector of the South to urban places. Earnings of domestic servants were understated in 1939 because of the receipt of income-in-kind payments (meals, sometimes lodging, and so on).

(2) Black earnings were relatively low in 1939, partly because of the high rate of unemployment throughout the 1930s. Black earnings rose sharply in World War II (1941–1945). The rate of increase in the men's black-to-white ratio has been slow but steady since the mid-1950s. During this period from 1955 to 1982, when real incomes were generally rising, the modest increases in the ratio have maintained roughly the same absolute difference in real earnings between blacks and whites.

(3) Blacks made relative gains between 1940 and 1960 in educational attainment and, probably, in other investments in human capital, such as health and access to better jobs by migration. In the 1960s and 1970s there were further gains in relative educational attainment and also in legal restraints on discrimination in employment.

(4) The ratio of women's earnings to men's among whites has been stable and reflects two counteracting trends: (a) more participation in the labor force by women and, associated with this, more accumulated work experience and ad-

vancement into higher occupations; (b) an increasing number of women who are new entrants or reentrants into the labor force, whose average years of experience are less than the average of the existing stock of women workers. Thus, (a) exerts a compositional effect that raises the ratio of women's earnings to men's earnings while (b) has the opposite effect [see Mallan (1982)].

The descriptive statistics presented in Tables 13.1–13.5 have shown two manifestations or definitions of economic discrimination, one dealing with incomes and another with wage rates, for three groups affected by discrimination: women, blacks, and Hispanics. The economic disparities are large and have persisted over time. The fundamental theoretical challenge is the presence and persistence of different wage rates for groups of workers for whom the assumption of equal productivity – or equal productive capacity – is maintained. The next section of the chapter surveys the economic theories that have been formulated in response to this challenge.

3. Theories of economic discrimination in the labor market

There is no shortage of theories to rationalize the existence of different wage rates for equally productive workers. What is scarce is a theory that is buttressed by empirical support. As discussed in the next section, the empirical work has seldom tested the theories. In this section I resort to informed opinion and speculative judgment about the plausibility and robustness of the theories.

Three theories of discrimination are found in the economic literature: (1) neoclassical, which include nonstochastic and stochastic versions, (2) institutional, and (3) Marxian. Only neoclassical theories, the basis for almost all the theoretical literature in the United States, will be examined in any detail. Marxian theory will not be examined, although certain components of this theory, such as exploitation, do appear in the neoclassical and institutional theories.

The neoclassical theory of discrimination is almost entirely a demand-side theory. The supply side of the labor market is effectively neutralized by the assumption that minority and majority groups of workers are equally productive (or have equal productive capacity) and have equal tastes for work. The demand side may be characterized by a competitive or monopolistic structure and by “exact” versus “stochastic” models. These characterizations define the taxonomy used below.⁸

⁸The taxonomy below, in subsections 3.1 and 3.2, of seven models was initially developed by Becker in his influential book that was published in 1957 and revised in 1971. I remind the readers of this point because “Becker's theory of discrimination” is often incorrectly identified with only one of his several models – that dealing with a competitive market and employers as agents of discrimination. The fact that I follow Becker's taxonomy in sections 3.1 and 3.2 should not be taken to mean that he would agree with my formulation of the models.

3.1. *Nonstochastic competitive neoclassical models: Discrimination by consumers*

Becker relabeled the abstract concept of "prejudice" into the economic concept of "tastes", and his operational definition of "tastes for discrimination" was that of a demand function; namely, a monetary offer for a good or service with, in this instance, a qualitative attribute (like race) that distinguishes it from another, otherwise identical, good or service. If the price of the labor service of the majority worker is p , then the prejudice or tastes for discrimination of a buyer are measured by an offer price, $p - d$, for the (otherwise identical) service of the minority worker. The term d is a measure of the buyer's tastes for discrimination. I use the small letter d to measure an individual agent's discrimination, and D will refer to marketwide discrimination.

Several advantages of the formulation are apparent. Discrimination has the appealing property of continuity, rather than being merely present or absent. It is potentially measurable, and the monetary units have an intuitive meaning to experts and laypersons alike—in contrast to various attitudinal scales ("like a lot" . . . "dislike a lot") that may or may not be scored numerically. There are explicit behavioral and even policy implications in the formulation. For example, a government subsidy to a minority-produced service could equalize the net price to consumers.

There are some disadvantages of the measure and some properties that may be either advantageous or disadvantageous depending on the question one is asking. No attention is paid to any pain or stigma felt by the victim. A lower price for one's services appears to capture the extent of victimization and to be on the same footing as a lower price owing to an inferior standard property of the good being sold. However, a black insurance salesman who offered the same policy as a white seller but sold and earned less because of customer prejudice might feel worse than if he received less because his policy offered less coverage or smaller settlements. Both price differentials could be the same, but only the former is viewed as an inequity and as a social problem.

Becker (1957, rev. 1971, p. 5) used the example of physical beauty as a qualitative attribute that leads to discrimination by demanders but is not ordinarily viewed as a social problem, either because beauty is considered legitimately productive—as it is in acting and modeling—or because discriminating in favor of this attribute is socially acceptable. On the one hand, whether discrimination in favor of an attribute is socially approved or disapproved is a datum to economists, just as we usually assume that preferences are given. Economists can still be useful if, after being informed of which attributes lead to socially disapproved discrimination, they are able to predict behavioral consequences and, ideally, suggest cost-effective remedies. On the other hand, inattention to the nonmonetary pain felt by the victim of certain types of discrimination will limit the economist's contribution to social welfare and policy analyses (to be discussed in the concluding section).

Economic analysts have generally concluded that consumer-based discrimination plays a minor role in the differences in average wages received by race and sex groups. The reasoning is as follows. Assume that black workers have the same distribution of productive skills as white workers and that consumers (who are predominantly white) are willing to pay a price, p , for a good produced by white workers. If, however, there is customer contact with the producers, the consumers consider the effective price for a good produced by black workers to be $p' = p + d$, where p is the cost of production and d is the monetary value of a white consumer's distaste for contact with black producers. (For convenience, assume temporarily that all white consumers have identical tastes.) Clearly, most goods and services are not produced with customer contact. Thus, consumers would not discriminate against, say, clothing or automobiles according to the color of the workers in clothing or automobile factories. For these goods the price would simply be p , regardless of the color of the workers.

Black workers, therefore, would specialize in the production of goods with no customer contact and, in so doing, avoid being paid a wage lower than that of an equally productive white worker, which would be the outcome if they competed with whites in, say, retail selling.⁹ If the concentration of black workers in industries with no customer contact were to depress wages in these jobs, then white workers in these jobs would move—horizontally by skill level—into jobs with customer contact until wages were equalized in the two sectors. Given that the number of black workers is small relative to the number of jobs that have customer contact, all black workers would be in jobs that have no customer contact. (Realistically, some would be in the jobs with customer contact that involve nondiscriminating customers, now recognizing that consumers have varying tastes regarding contact with black workers.) The result is some degree of job segregation but no group difference in prices for labor services.

The assumptions that lead to this outcome are sufficiently plausible that consumer-based discrimination has not been assigned an important role. The market measure of discrimination, D , equals zero, even though consumers are prejudiced and job assignments among workers are affected. Thus, Becker's formulation provides the useful distinction between an i th individual agent's tastes for discrimination (with potentially varying d_i 's) and market discrimination, which is an aggregate that is not the sum of its parts; here, $D = \bar{p}_{\text{maj}} - \bar{p}_{\text{min}}$. Discrimination, D , disappears even though $\sum d_i > 0$, simply because workers, in their quest to maximize their utility, will move and bring about some degree of segregation.

An outcome in which segregation reduces or eliminates market discrimination occurs in several versions of Becker's model. For this reason, Welch (1975) called Becker's theory a theory of segregation, not discrimination. Welch's point is

⁹Specialization, which is here associated with segregation, has been rigorously analyzed as a means for attaining nondiscriminatory outcomes in terms of factor payments by Stiglitz (1973, 1974).

partly semantic, but his insight is useful and may be explained briefly as follows. The source of market discrimination in Becker's model is on the demand side – the willingness of an economic agent to pay to avoid contact with members of a specific group. In a competitive model there are many employers and free mobility among economic agents, so competition enables segregation to satisfy this demand costlessly. The model assumes that mobility is costless (or nearly costless), especially in the long run.

Segregation is, therefore, a means for eliminating market discrimination, but it is not the only means. Collective action to offset the effects of discriminatory tastes or changes in those tastes can be accomplished without seriously restricting competition in markets. Indeed, common sense and casual observation indicate that an integrated society is generally more competitive. It is tempting to point to the Republic of South Africa to illustrate that segregation is not a sufficient condition to eliminate discrimination, but this country's experience is inappropriate for illustrating a competitive model. Lundahl and Wadensjö (1984, pp. 209–260) explain how a century-long pattern of private and governmental collusive arrangements have restricted competitive forces in the South African economy, with the undisguised purpose of concentrating wealth and power in the hands of the white population.

3.1.1. Discrimination by workers

Assume all workers have the same skill level. If all majority workers (whites) are prejudiced against minority workers (blacks), we may assume a white worker's wage demand for working with other white workers is w , and his wage demand for working with black workers is $(w + d)$. Clearly, employers of white workers would employ segregated work forces to pay the lower wage. Equally skilled black workers would also receive w as a consequence of competition among employers, mobility by workers, and the previously established sterilization of consumers' preferences. Integrated work forces could exist among unprejudiced white and black workers, so the worst case is when all white workers have tastes against working with black workers. But even the worst case yields only segregation among workers, not discrimination as defined by $\bar{w}_{\text{maj}} > \bar{w}_{\text{min}}$.

One could postulate various impediments to competition. For example, perhaps segregation will not permit equal wages because the black workers are too few to allow economies of scale in production, recognizing that their numbers must staff all skill levels. Rebuttal: Aside from examining the structure and technology of industrial organization to determine the plausibility of this, we should recognize the flexibility in large-scale organizations to use compartments, work in shifts, form subgroups, provide on-the-job training, and so on to achieve "effective" segregation of the workers. Remember, segregation is cost minimizing when white workers are the discriminatory agents.

Another example: Assume that black workers migrate into a region populated exclusively by prejudiced white workers. Efficient, segregated firms might take a long time to become established. Hiring and training workers entails fixed costs and, as Arrow has analyzed, these costs will retard any attempt by a firm to hire an all-black work force [Arrow (1973, pp. 20–23)]. A rebuttal should not be required because the example, although empirically relevant and interesting, should lead to a long-run equilibrium in which the work force is segregated.

Another example: Let skills vary among workers and assume that black workers have a legacy of low skills upon entering the labor market. White workers with equally low skills receive w' . Assume the technology of efficient production requires that low-skilled workers combine with complementary high-skilled workers, all of the latter being prejudiced white workers. Black workers must then receive $w'_{\min} < w'_{\text{maj}}$ to compensate for (offset) the high labor costs they “impose” on their complementary factor of production – the white skilled workers. Rebuttal: Some black workers would have a particularly strong incentive to become skilled. Those who match the skilled white workers in innate ability would not only have the incentive to seek the normal (i.e. white workers’) rate of return on a skill investment, but they could earn extra profits by working with low-skilled fellow black workers, because the discrimination tax, d , will not apply to them. To see the incentives involved, we can imagine that these tax savings could be shared among both skill levels of black workers and their employer, and any one of these agents would have an incentive to initiate this process. Eventually, as more black workers become skilled, the underlying source of the $(w'_{\text{maj}} - w'_{\min})$ gap withers away. Again, this scenario could take a long time, and it may be empirically interesting. Finally, I argue in the next section that complementary skilled white workers correspond to employers as agents of discrimination, so the conclusions about employers also apply to complementary skilled workers.

3.1.2. *Discrimination by employers*

Two versions of employer-based discrimination in competitive markets were advanced by Becker in his analysis of racial discrimination. The first, hypothetical and pedagogic, assumed that employers all have the same prejudice against black workers (or in favor of white workers), so a uniform lower demand for black workers sets their market wage at $(w - D)$. Thus, the white workers’ wages and their monetary labor costs are higher. Competition in the product market requires a uniform product price, but this can be achieved by the differential in money labor costs being compensated by a differential in money profits, which, in turn, is compensated by a differential in psychic benefits (or psychic costs – the difference depending on whether one emphasized the employer’s psychic benefit from employing a white worker or the psychic cost from employing a black

worker). The psychic and money forms of profits (or employer compensation) offset one another in equilibrium.

Other analysts suggested modifications of this model. Arrow (1972) obtained useful insights from an assumption that the employer's discriminatory tastes were an increasing function of the ratio of black-to-white employees, rather than being a constant that was independent of the racial composition in the firm. Arrow (1972, p. 89), Marshall (1974, p. 853), and Thurow (1975, p. 162) suggested that distaste may depend on "social distance" rather than "physical distance". If true, this would make empirical measurements complicated. For example, an employer's d might be zero for janitors but have a large negative value for professional employees. Indeed, if the owners of capital have little or no contact of any kind with the employees, the model would require that the discriminatory role shifts from employers-as-capitalists to their agents, such as managers, supervisors, foremen, or even skilled workers—all of whom are assumed to be prejudiced white persons. These interpretations of employer discrimination add realism to the model, but they do not negate Becker's central point, which was the establishment of an equilibrium differential in favor of white workers.

In a second version of Becker's model of employer discrimination in a competitive economy, tastes among employers were permitted to vary. Consider, first, the special case of just two values of d_i : low, d_1 , and high, d_2 . Clearly, employers with the lower value, d_1 , would hire all the black workers. (I will temporarily assume that there are enough d_1 employers to hire all the black workers.) The market wage differential between white and black workers under this regime would be $D_1 = d_1$, a smaller differential than the average: $(N_1 d_1 + N_2 d_2)/(N_1 + N_2)$, where N_1 and N_2 are the numbers of employers in the two categories. Indeed, the size of d_2 is irrelevant.

Becker's insight from this model is that black workers generally benefit by a dispersion in d_i . A wider spread in the distribution of d_i could only narrow the wage gap, assuming some of the increased variance stretches the lower tail of the distribution and lowers the value of the d_i of the employer with the highest d_i required to hire all black workers. Intuitively, the upper tail is irrelevant in the setting of D because the employers with larger tastes for discrimination, $d_3 > d_2$, $d_4 > d_3$, and so on, do not bid for minority workers and they have no incentive to pay more than the existing w for majority workers. In contrast, a widening spread in the lower tail means that the new employers, with tastes $d_0 < d_1$, would now hire all the black workers. They increase the demand for black workers, and the market differential in white and black wages becomes $D_0 < D_1$.

Two plausible extensions of the dispersion effect, as just described, will tend to eliminate market discrimination entirely.

(1) First, the lowest value of d , call this d_0 , would determine the market wage differential, even if only a small number of employers—in the limit, one per

product per market – had a value of d_i as low as d_0 . Clearly, this employer would earn extra profits by hiring minority workers, benefitting monetarily from the lower wage they receive while escaping all or some of the psychic costs that would be experienced by employers with higher d_i 's. Total profits could be increased by cutting prices and hiring more black workers and expanding production. Employers with $d_i > d_0$ would, correspondingly, lose business and curtail production, thereby decreasing the demand for white workers. The impersonal operation of the capital market would ensure an inflow of investment to the high-profit firms. Assuming long-run constant costs, the stopping point would be reached only when all black workers and equally paid white workers are employed by the d_0 employer(s) – perhaps in newly constructed plants, each of optimal size. White workers would lose the wage advantage they had received from discriminating employers.

(2) Second, D_0 would become zero. There are several routes by which the market should uncover one or more cost-minimizing employers (per product, per market) with $d_0 = 0$. Some white employers might be unprejudiced. Blacks could become employers. Capital owners, like consumers, tend to be remote from contact with employees, so their d_i 's would tend to be effectively zero. (Of course, this shifts the cost-minimizing problem to that of finding managers with low d_i 's.) Indeed, consumers as well as investors would have precisely these incentives of finding managers and other forms of complementary employees whose $d_0 = 0$.

In a phrase, competitive market forces, still assuming constant costs, tend to drive D toward zero. Arrow, in his analysis and reformulation of Becker's model of employer discrimination, arrived at just this conclusion: "Only the least discriminatory firms survive. Indeed, if there were any firms which did not discriminate at all, these would be the only ones to survive the competitive struggle" [Arrow (1973, p. 10)]. And, "It [Becker's model of employer discrimination] predicts the absence of the phenomenon it was designed to explain" [Arrow (1972, p. 192)].

Becker, in an article on discrimination written for the *International Encyclopedia of the Social Sciences* and published 11 years after his book, did not reach this conclusion.

A few of the more extreme nineteenth-century advocates of a competitive market economy believed that eventually its extension and development would eliminate most economic discrimination.... Unfortunately, this has not yet taken place; discrimination exists, and at times even flourishes, in competitive economies, the position of Negroes in the United States being a clear example [Becker (1968, p. 210)].

Becker's disagreement with the previous scenario of the workings of competition is based on his view that the assumption of constant costs for a firm, even in

the long run, is a polar case and not one to be accepted generally [Becker (1957, rev. ed. 1971, pp. 44–45)]. Entrepreneurial skill is an example that is sometimes suggested for a factor of production that may be inelastically supplied, even in the long run. Thus, one's judgment about the number of nondiscriminating firms that are in or that might enter the market, about the generality of entrepreneurial skills, and about the long-run elasticity of other factors all enter into one's judgment about the persistence of a discriminating cost differential in the long run under competitive conditions.

What if discrimination is redefined as nepotism and $d = d_b < 0$ is replaced by a term $d_w > 0$, now adding subscripts to distinguish discrimination against blacks from nepotism in favor of whites? This specification is examined by Goldberg (1982), who finds that a long-run differential wage advantage in favor of whites is sustained under competitive conditions. The result, which had been previously advanced and then downplayed by Arrow (1972), is correct, but in my judgment the model is not realistic.¹⁰

My argument begins with the observation that when a positive d_w – Goldberg's nepotism – replaces a negative d_b – Becker's discrimination – the intention is to view the tastes for whites as more than a euphemism for expressing a preference not to be associated with blacks. This intention is clarified by a dictionary definition of nepotism: "favoritism shown to one's nephews and other relatives; bestowal of patronage by reason of relationship rather than merit". As defined, nepotism is indeed real. Let us assume that only the "uncle-employers" receive nonpecuniary utility from the employment relation. Consider two cases of wage payment. In Case 1 the wage rate of "nephews" and all other workers is the same, and nephews are merely sorted into jobs where their uncles are employers. Alternatively, in Case 2 the uncles share all or some of their utility rents with their nephews by paying them a higher than competitive wage. In Case 1 the uncles earn extra rewards (profits plus utility), but they have no incentive to expand production, which would (assuming constant costs) threaten other firms, because the supply of nephews is sharply limited. In Case 2 the uncle-employers earn lower profits, but their total utility can easily be high enough to ensure their survival as employers.

Case 2 shows, therefore, that the dictionary definition of nepotism can coexist with the economic definition of nepotism, according to which nephews receive a higher wage than equally productive nonnephews (all other workers). However, to transfer this scenario of nepotism to one in which all white workers, who

¹⁰ See Arrow (1972, pp. 91, 192). After pointing to nepotism as a source of a sustained wage differential in favor of whites, even in the absence of any differential based on tastes against blacks, Arrow commented: "But it is reasonable to postulate that any preference a firm might have for the hiring of whites per se arises as an offset to the presence of disliked blacks. That is, for a firm that has no black employees, $d_w = 0$." Furthermore, "for a firm that does not discriminate against blacks, there will also be no reason to pay anything extra for white employees" (p. 192).

constitute 85 percent of the labor force, are the equivalent of nephews (beneficiaries of nepotism) seems unrealistic. Throughout this survey, therefore, discrimination against a minority group will be viewed as the operative force.

3.2. Nonstochastic monopolistic neoclassical models: Product monopoly

A monopoly has two characteristics that permit long-run discrimination: first, a definitional uniformity in tastes, since there is only one employer; second, above-competitive profits. The former allows a d_i that will not become irrelevant because of competition, and the latter allows the sacrifice in money profits – in exchange for the psychic benefits from discrimination. Nevertheless, there are several influences in the economy at large that constrain or even eliminate the power of one or a few monopolies to sustain market discrimination.

Monopoly power in the product market does not imply monopoly power in the labor market. If the monopoly firm cannot affect wages in the labor market, it would not pay a higher wage than w to hire majority workers, nor could it pay a lower wage than w to hire minority workers. In other words, the monopoly would not be the source of discrimination, although it, like other firms with a positive d_i , would employ a segregated, all-majority work force. Were the monopoly to behave irrationally and pay higher wages to majority workers, it would create incentives for a “takeover” by investors and managers with zero d_i 's. Indeed, Alchian and Kessel (1962) advanced the view that even where monopolists affect wages in their labor market, they would be unlikely to sacrifice money profits permanently by a policy of (racial) discrimination, because profit-maximizing investors would buy them out.

But why do monopolistic enterprises discriminate...more...? One would expect that those who have a taste for discrimination...would naturally gravitate to those economic activities that, for purely pecuniary reasons, do not employ Negroes. Free choice of economic activities implies a distribution of resources that would minimize the costs of satisfying tastes for discrimination (p. 161).

Alchian and Kessel pointed out that a regulated monopolist or a government monopoly, which was constrained *not* to maximize profits, could indulge its tastes for discrimination at no loss in profits and, therefore, offer no incentive for a “takeover”. Such firms could, for example, engage in nepotism and consume other nonpecuniary benefits at no cost in forgone profits, and if there were enough such firms, they could at least contribute to a marketwide discrimination differential.

It is useful to keep in mind two empirical characteristics of monopolies – now using the term as shorthand for a firm that produces a “large” share of the market. First, monopolies tend to be larger, more capital-intensive, and more

likely to be unionized than the average firm. Because of this, they may pay higher wages to attract specialized skills and to ensure lower turnover. Among the workers who apply for jobs at these monopoly firms, majority workers may be the more skilled, as a result of previous discrimination from various channels. The resulting combination of hiring relatively more majority workers and paying higher wages may not be discriminatory; that is, it may be consistent with a $d_i = 0$ for the monopolist. In principle, a properly specified Model (I) would permit testing whether the firm really discriminated among equally skilled applicants, minority and majority.

Second, along with size and wealth, monopolies are often also publicly prominent. They tend to be sensitive to public relations and to their "image". In the past this sensitivity could have served to reinforce discrimination, because government and other wielders of power in the community may have been prejudiced and have influenced the monopoly. Today, our laws and professed public sentiments are against discrimination or, if neutral, condone organized pressures from minority groups on the monopolies. These forces would, if present, tend to lower the effective d_i of the monopolist below the average among employers.

In summary, monopoly firms, particularly regulated monopolies, are in theory capable of exerting some sustained discrimination in labor markets. There are, however, reasons for doubting that monopoly is a major source of marketwide discrimination.

3.2.1. Monopsony firms in labor markets

The classic case of the exploitation of labor in neoclassical economics arises under conditions of monopsony. Workers are captive in a market where there is only one employer, or where a group of employers collude and act as one buyer. Monopsony represents a rare area of common ground between neoclassical and Marxian models of the labor market.¹¹

The model is well known: a single buyer of labor faces an upward-sloping supply curve of labor; equates the value of labor's marginal product (VMP) and its (rising) marginal cost; hires less labor than if the same demand for labor were generated by many competing firms; pays labor its supply (offer) price, which is lower than the price (wage) needed to induce the larger supply under competitive demand conditions; and retains the positive differential between the VMP and the wage as profit. Where two factors of production are supplied and demanded, the exploitation (measured by $[VMP - w]/w$) will be greater for the factor whose labor supply is the more inelastic. These propositions, which were presented by

¹¹See Lundahl and Wadensjö (1984, pp. 49–52) for a further analysis of monopsony models of labor market discrimination and for their critique of neo-Marxist, or radical, theories as a subset of monopsony. See Cain (1976) for a brief discussion of radical theories of the labor market, including the analysis of discrimination by radical theories.

Joan Robinson (1934, pp. 301–304), provide a consistent model for discrimination simply by postulating a more inelastic supply curve of labor for minority workers. A modern application of this model is by Madden (1973).

Empirical support for the prevalence of monopsony and lower-than-competitive wages is limited [see Bunting (1962)]. Labor markets that are “one-industry towns” are increasingly uncommon, mainly because a large fraction of the population lives in larger urban places and because the automobile has greatly expanded the geographic boundaries of the labor market. Information about wage rates in geographically dispersed markets is available and only those workers “on the margin” of moving need to move to equalize wages for workers of comparable skills. Therefore, the long-run acceptance by workers of below-competitive wages presupposes a degree of immobility that is hard to accept. No doubt there are some workers who are trapped by a combination of industry-specific skills and a decline in the number of firms competing for their skills, and who suffer long-lasting exploitation. But these are not conditions that generalize to the entire labor market.

Because monopsony seems to have a limited application, it does not appear worthwhile to examine more closely the requisite proposition that the supply curve of minority workers is less elastic than the supply curve of majority workers. However, two brief points may be useful. First, if differences in the supply curve identify (in the econometric sense) a difference in exploitation, we need to satisfy ourselves that the underlying sources of this difference in supply curves are not also reasons why the workers’ wages differ.

Second, regarding gender discrimination, there is a good deal of empirical evidence and theoretical support for the finding of a greater elasticity for the supply curve of women’s labor than of men’s labor. To be sure, this larger elasticity refers to the market, not to individual firms, but as a firm (or group of firms) becomes monopsonistic then the distinction between the supply of a factor to the labor market and the supply of a factor to the (monopsonist) firm tends to disappear. Thus, the larger labor supply elasticity of women in the labor market as a whole implies a larger elasticity to a monopsonist, and this is the opposite of the requisite condition for the exploitation of women relative to men. Again, there may be particular circumstances when this generalization does not hold. Nurses are sometimes used as an example of an occupation that faces a monopsony-employer in the form of one or a few hospitals.

3.2.2. *Labor unions as monopolies*

In Becker’s model of discrimination, white workers’ prejudice against black workers was not a sufficient condition to sustain a discriminatory wage differential. However, by forming a monopoly in the sale of labor to employers, white workers could enforce their tastes *and* raise their wage above the competitive level. Moreover, unlike monopsony, labor unions are widespread, supported by

laws and community approval, and have been shown in many studies to have raised wages for their members above competitive levels.

Given that the union secures monopoly rents, some method of restricting entry is a necessary first step in maintaining these rents. Many analysts have pointed to the discriminatory tastes of the members as a criterion for inclusion and exclusion. Kessel (1958) added the argument that this criterion will also be useful in a second step in maintaining the rents; namely, in policing the existing members to honor the union contract, even though it would often be in their private interest to "cheat" by, say, working more for a slightly lower wage. Kessel argued that ethnic homogeneity among the members facilitates a mutual agreement to collude, making unnecessary those stronger sanctions that might be illegal or incur community disapproval. Finally, institutional research, while divided about the overall discriminatory impact of unions, documents many cases of discrimination by unions [Gould (1977), Hill (1977), Marshall (1965), Ross (1948), Northrup (1944), among many others]. Thus, the a priori case for unions as a source for labor market discrimination appears substantial.

There are, however, a number of counterarguments. First, unions have never organized a majority of the labor force in the United States, and before 1940 there were few periods during which more than 15 percent of the work force was covered by collective bargaining contracts. The wage gap between blacks and whites was larger in the pre-1940 period, although this fact by itself does not provide direct evidence on the influence of unions on the wage gap. In 1977, only around 25 percent of the labor force were union members or were covered by collective bargaining contracts [U.S. Department of Labor (1979)].

Second, membership in unions is more common among blue-collar workers, which points to a disproportionate representation among men and blacks, although within the blue-collar ranks membership is more common among skilled occupations, which points to a greater representation among white men. A larger proportion of black men were members of unions in 1977 than were white men [U.S. Department of Labor (1979)].

It is noteworthy that the few industries and occupations where unions have grown in recent years – governments, teaching, hospitals – are disproportionately composed of women or blacks. Ashenfelter (1972), whose study will be examined in the next section, concluded that the white-black wage gap among men was actually narrowed by unions as of the mid-1960s. The male-female gap was slightly widened. His study is persuasive that labor monopoly, despite many individual cases of discrimination by unions, is not a major source for the observed discriminatory differentials.

3.2.3. Government as a monopolist

Governments are universally monopolists in certain functions, such as providing for national defense, police and fire-fighting, and mail services, and, most

importantly, as law-maker. With their power to tax and to punish, governments possess more potential monopoly power than firms and unions, although the collaboration between government and private agents may make it difficult to isolate the source of power. Moreover, governments, unlike private monopolies, need not be and seldom are guided by profit maximization goals. Granting that the majority group controls the government, there is no analytical challenge to demonstrating a theoretical case for discrimination based on government behavior. Malcolm Ross, the director of the Fair Employment Practices Commission during the 1940s, provides an example of a government law that, if it did not impose wage discrimination against blacks, at least impeded its demise. Ross's example also illustrates one expert's skepticism about the "physical-distance" theory of discrimination.

White and Negro workers are now [1948] and have been for decades under the same plant roofs in the South. It is not the working associations to which the whites object. It is the sharing of skilled wage rates.... South Carolina...refuses by law to permit skilled Negro textile workers in the same plants with whites. But that state statute (probably unconstitutional) does permit Negro janitors and charwomen to work under the same textile plant roofs as whites. What would you say, then, that that South Carolina law is protecting - white workers from association with Negroes, or white jobs at the looms at white wages? [Ross (1948, p. 307)].

The scope, history, and literature of the government's influence on labor market discrimination are far too extensive to survey in this paper. Some discussion about government policies is reserved for the final section.

In this chapter I generally assume that government agencies do not have pervasive monopoly power regarding labor market discrimination, and that, historically, their interventions in the market have had many, but more or less offsetting, effects. In recent decades the intention of government policies has been to reduce discrimination against minorities, but the analysis is complicated by the claims that some actions, despite the beneficial intentions, turn out to worsen the problem. This criticism is frequently made about minimum wage and equal pay legislation. One fact and two theoretical-empirical points set the stage for this criticism.

Fact. The minority group is disproportionately represented in the lower tail of the distribution of productive skills, not because of an innate inferiority but because of a legacy of past inequities and prelabor-market discrimination.

Theoretical Case 1. Model (I) applies with the competitive result that minority workers receive an average pay equal to their average abilities (defined by X). However, the minimum-standards law truncates (from below) the distribution of X 's among the work force, and relatively more members of the minority group are unemployed from the jobs covered by the legislation. Over the full distribution of the work force, minority workers are worse off, either because of their

excess unemployment or because those disemployed from covered jobs are crowded into lower-paying jobs in the uncovered sector. Note that this case does not require any tastes for discrimination, although they would exacerbate the minorities' disadvantage (see Case 2).

Theoretical Case 2. Model (I) applies and majority workers receive a higher wage, conditioned on X , implying $A > 0$ and the existence of market discrimination. A minimum wage, w , can impede the competitive forces that encourage hiring lower-wage minority workers. Employers who might hire minority workers at a lower wage, $w - D$, are prevented from doing so. Case 2 does not require minorities to be concentrated in the lower half of the productivity distribution, but lower-skilled minority workers face the highest risk of being without a job. They are also prevented from competing for jobs that offer general on-the-job training by bidding for them with lower starting wages.

To illustrate either of the two cases, consider the following historical event, described by Ross (1948).

During the First World War the Southern [railway] carriers lost a serious number of skilled workers to the services and munitions plants. In order to make it attractive to Negro workers to stay on the job, Secretary McAdoo as wartime transportation chief ruled that Negro railwaymen should receive the same pay as whites for the same work. This 1918 move was called "a simple act of justice," and so it was, although the far [long-run] results were anything but just.

Forced to pay them the same wages as whites, the carriers lost interest in Negroes as a cheap labor supply. The white workers, for their part, began to covet the better Negro jobs. The McAdoo ruling had laid the foundation for a coalition between the carriers and the unions against Negroes in firemen's and other high-bracket positions (p. 119).

Blacks were driven out from these positions, but as Ross makes clear, the government ruling was only one part of the causal chain. Also contributing to the outcome were employer and worker prejudices, a quasi-monopolistic industry, the antiblack environment of the South, and a labor union.

3.3. Stochastic neoclassical models and statistical discrimination

The theoretical challenge developed in the preceding discussion of neoclassical models is to rationalize unequal pay to groups of workers who are equally productive. The comparison between groups was intended to allow within-group individual deviations from the equality between productivity and pay, which is necessary if the model of pricing is to apply to the real world. However, this stochastic feature was suppressed throughout the discussion, because the use of

average values of the wages for comparisons between groups made the models equivalent to exact or nonstochastic models.

Attention to a stochastic model of wage determination, in which the worker's value to the employer is not known with certainty, offers several new insights, and more possibilities for sustained (or, at least, long-lived) group discrimination. Whether these theories are more or less persuasive than any of the others is a matter for judgment and empirical study.

Phelps (1972), Arrow (1972, 1973), and McCall (1972) were early authors. It is convenient to analyze the following model of wage determination, which is due to Phelps. Let q_i be the i th worker's true productivity, which is unknown to the employer, who must rely on some observed but imperfect indicator, y_i . The indicator may be a test score or a variable, like years of schooling, that has a more direct connection to productivity. The notation and details of the model below are shown in Aigner and Cain (1977), along with citations to various authors and statistical references.

In a simple specification that brings out the main conclusions of the approach, the relation between y and q (subscripts dropped) is

$$y = q + u, \quad (1)$$

with $E(u) = C(q, u) = 0$, $E(y) = E(q) = \alpha$, $V(u) = \sigma_u^2$, using the familiar symbols for expectation, covariance, and variance. By assuming q and u are joint-normally distributed as well as uncorrelated, we may specify a linear regression function for the reverse regression:

$$q = \alpha(1 - \gamma) + \gamma y + e, \quad (2)$$

with e a well-behaved disturbance. Here γ is the coefficient of determination (r^2) between q and y ; thus, $0 \leq \gamma \leq 1$, and γ measures the "reliability" of y as a measure of q .

Assuming employers pay workers according to their expected productivity, then

$$w = E(q|y) = \alpha(1 - \gamma) + \gamma y. \quad (3)$$

Equations (2) and (3) reveal the obvious point that *individual* discrimination, defined as unequal pay for equally productive workers, is inevitable, given the error component, e . In contrast, *group* discrimination does not follow from this model precisely because e is considered random and has an expected value of zero for minority and majority groups.

Letting subscripts 0 and 1 refer to minority and majority groups, eq. (3) may be applied to each group. Assume temporarily that the minority and majority

groups have the same mean true productivity: $\alpha = \alpha_0 = \alpha_1$, and that we compare workers with the same y -score. If we further assume that $V(q)$ is the same for both groups but that $V_0(u) > V_1(u)$, reasoning that the test instrument is more unreliable for the minority group, then $\gamma_1 > \gamma_0$, and we have

$$w_1 - w_0 = (y - \alpha)(\gamma_1 - \gamma_0). \quad (4)$$

Accordingly, for a given y -score [roughly corresponding to "holding X constant" in Model (I)], majority workers receive a higher wage than minority workers for y -scores above the mean, α , and lower wages for y -scores below the mean. Thus, group discrimination, defined by $E(w_1 - w_0) > 0$, is not present.

Clearly, postulating a lower α for minority workers would lead to their being paid a lower wage, but a lower wage for a given y -score, assuming y is a valid indicator of productivity (about which, see below), would not imply economic discrimination for the group because, on average, the minority and majority workers continue to be paid in accordance with their average productivity. In Figure 13.1, parts (a) and (b) show two cases for unequal α 's. In part (a), where $\gamma_1 = \gamma_0$, the difference $\alpha_1 - \alpha_0$ is evenly distributed across all y -scores. In part (b), where $\gamma_1 > \gamma_0$, we see that the minority workers with high y -scores who are paid "too little" relative to majority workers with the same y -scores are balanced by the low-scoring minority workers who are paid "too much", relative to majority workers with the same low y -scores. As drawn in part (b), minority workers with y -scores below y' receive relatively higher wages than majority workers with the same y -score.

Nevertheless, a number of economists have claimed that this model reveals, and offers an explanation for, group discrimination. Let us examine two applications of the model. Only the second shows discrimination that is consistent with the definition adopted in this chapter.

3.3.1. Statistical discrimination, but spurious group economic discrimination

Thurow (1975) is one of many economists who use the term "statistical discrimination", when there is presumptively no economic discrimination. In the following example, Thurow accepts the facts of (a) a higher probability of market work by men compared with women and (b) the benefit to an employer of the higher probability. He then says:

Any employer faced with these differences in work probabilities will practice statistical discrimination even though there are millions of women who will be in the full-time paid labor force for their entire lifetimes. *Ex ante*, he cannot tell which women will be lifetime year-around full-time employees and which women will leave the labor force or become part-time employees. Because the employer provides on-the-job-training, he will want to invest in those who are more likely to stay in the full-time labor force. If he provides training to

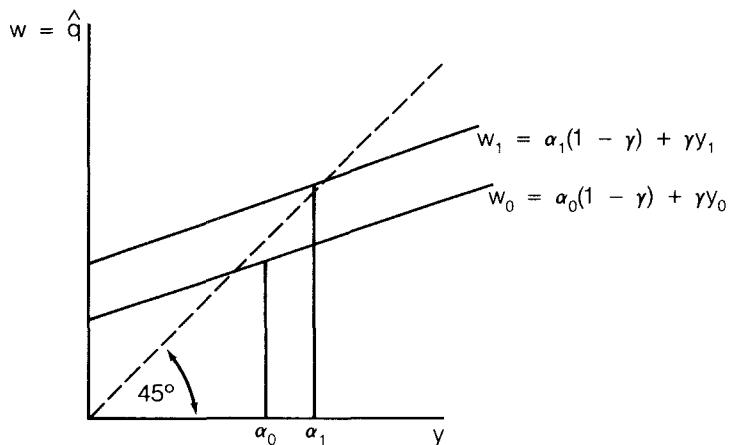
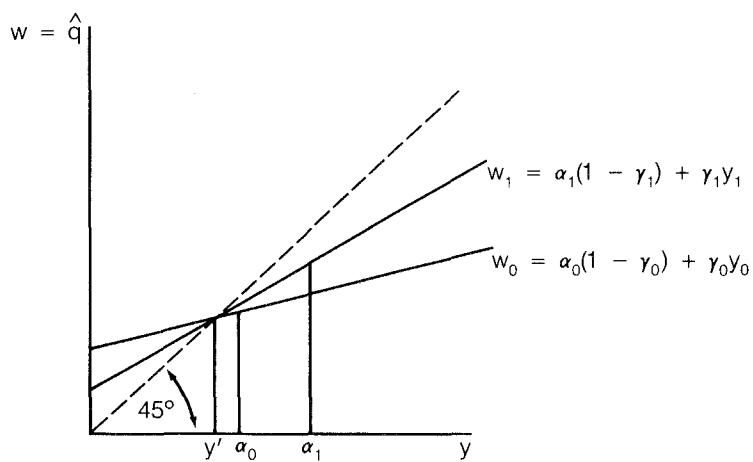
(a) $\gamma_1 = \gamma_0$; $\alpha_1 > \alpha_0$ (b) $\gamma_1 > \gamma_0$; $\alpha_1 > \alpha_0$ 

Figure 13.1. Predicted value of productivity (q) by indicator (y) for majority (1) and minority (0) workers.

women, he is less likely to be able to recoup his investment.... The woman who will participate in the paid labor force her entire lifetime is being treated unfairly.... The net impact is discrimination against women as a group and as individuals even though there is not a basic taste for discrimination against women (p. 178).

Two points should show why this example does not imply group economic discrimination. First, Thurow correctly indicates that the women who will participate in the labor force their entire lifetimes are being treated unfairly and will be underpaid. The employers cannot know an individual's future, and they will base their wage offer partly on $\alpha_0(1 - \gamma)$ —that is, partly on the known average for all women. But this is only half the story. Women who will participate for only the briefest period will be overpaid. As before, the employer, not knowing these women's true low probability of working, will rely upon the average for all women and overpay them. On average, the over- and underpayments tend to cancel out. Whether the resulting average is equal to the average for men will depend, as the next two paragraphs suggest, on whether the gender difference (here, a commitment to full-time work) is related to productivity.

Second, suppose all the workers are the same gender, that the two groups under study are persons with a college education and persons with less than a college education, and that the former have a higher probability of working on average. Thurow's entire passage could stand intact with the phrase, "persons with less than a college education", substituted for "women". Most analysts would agree that Thurow's case for group discrimination, even with the less-educated group earning less on average, loses its plausibility with this substitution.

Thurow's example inadvertently raises another interesting issue. The y -indicator in the stochastic model of wage determination is assumed to be unbiased on average, even though its reliability may differ for minority and majority workers. When, however, the y -indicator reflects discrimination, the model is no longer appropriate for an explanation of discrimination. In Thurow's example, the probability of working is, or could be, a reflection of discrimination. Clearly, if women or other minorities are discriminated against by not being employed, it is unsatisfactory to use the low probability of employment as an *explanation* for discrimination in the form, say, of lower wages or some other labor market outcome. This point will be discussed in Section 4, and here it serves to remind us that the choice of a y -indicator is not innocent.

3.3.2. Statistical discrimination and actual group discrimination

The discussion of the stochastic model up to now has not allowed the unreliability of the indicator to influence the average wage. Aigner and Cain (1977) stipulated risk aversion in the employer's utility (or profit) function and ra-

tionalized a lower average wage payment to minority workers as compensation for this undesired unreliability. A more convincing rationalization was suggested by Rothschild and Stiglitz (1982), who specified a production function that depended directly on matching the worker's q with a job assignment. In particular, both undermatching and overmatching were inefficient, so the expected output, not merely its variance, depended on matching.

Either formulation may be viewed as redefining the productivity of workers to include both the workers' physical productivity and the information workers convey about it. Does rewarding a group for their better information constitute economic discrimination against the group with less complete information? Perhaps the answer depends on the fairness of the testing system and, like the issue of the existing technology (see pp. 698–699), on how costly it is to change, and on whether its existing inadequacies for minorities reflect some market failure. The important role of the government in educating, training, certifying, and licensing workers suggests that improvements in testing minority workers may be a public good. (In fact, improvements in testing all workers may be a public good, but I focus here on discrimination between groups.)

If wage differentials are large merely because of differential test reliability, then both minority workers and employers have incentives to improve the tests and reduce this impediment to transactions. If, as is sometimes reasonable to assume, the worker knows his or her own abilities, a low-cost private-exchange method of minimizing this impediment is for workers to offer a trial period of employment to demonstrate their true productivity. The cost to the worker is a low wage during the trial period, but the benefits are higher earnings throughout the worker's subsequent career.

A trial period of working is also a device for minimizing the private and social costs of "signaling", as the term has come to be used following Spence (1973). Using Spence's model, we may assume that the test or y -score (a) has no value other than to indicate (signal) the worker's productivity, (b) is costly to obtain (as when the signal consists of an educational degree), and (c) is more costly to obtain for less productive (less able) workers. These assumptions imply that workers will choose whether to invest in the signal on the basis of their knowledge of their ability and on whether the extra pay the signal earns for them will justify its investment costs. Employers adapt to this maximizing behavior of workers by believing the signals and making their wage offers accordingly.

In Spence's model there is no guarantee that the equilibrium allocation of signaling investments among workers and, correspondingly, of workers to jobs is socially efficient, because only a "justifying" benefit/cost structure and not an "optimizing" one is required for an equilibrium. There is a tendency for too much investment in signaling; that is, Pareto-optimality could be achieved with less. Without an optimizing equilibrium there is no guarantee of a unique equilibrium. With multiple equilibria, the door is open for a benefit-cost structure that is unfavorable to a minority group compared to the majority group.

How robust is this discriminatory equilibrium? Even if one did not have faith that the competitive market would facilitate efficient signaling instruments and institutions, there remains the previously mentioned method of trial work periods based on deals struck between individual workers and employers. The strategy assumes that if the workers know enough about their ability to choose whether to invest in the signal, then they can use this knowledge to offer to work for the employer for a trial period. The strategy is better able to eliminate the Spence type of discrimination than it is to eliminate the Spence type of social inefficiency. Discrimination is eliminated if the cost to minority workers of the trial period is no higher than the cost of the majority worker's signal, even though these costs may still be higher than the socially efficient level.¹²

A recent paper by Lundberg and Startz (1983) uses certain features of both the Phelps model of unreliable indicators and the Spence theory of signaling. They derive a market failure in investment, although in contrast to Spence, too little investment occurs rather than too much. Their argument may be conveyed by reference to a commodity. Assume the commodity is produced at less quality than would be optimal, because the information about its quality cannot be conveyed perfectly. Specifically, the quality improvement could be produced at a cost that is less than the benefit, if only the quality improvement were accurately conveyed. Because the quality is imperfectly measured, however, consumers will discount the quality signal and will pay less than the costs of the optimal amount of the quality improvement. The situation is the same as in Model (3) above: employers pay γ (< 1) for a unit more of y , instead of paying a full unit more as they would if y were a perfect measure of q .

Lundberg and Startz apply this argument to two groups of workers, minority and majority, and show that a less reliable signal for minority workers will lead them to underinvest relative to majority workers. Their general conclusion of underinvestment is opposite to that of Spence because of their contrasting assumptions about the benefits and costs of the investment. For Spence all or part of the benefits were merely in "signaling", whereas for Lundberg and Startz all of the benefits are in the form of enhanced productivity. For Spence, the costs of the investment varied inversely with the productive ability of the worker; for Lundberg and Startz, the costs are invariant with respect to the productive ability of the worker. Apparently, a proper mixture of the two sets of assumptions could yield optimal investment. Both models face the criticism that the employer's uncertainty about the productivity of workers may be inexpensively reduced by observing the worker's on-the-job performance.

¹²An illustration of the adaption of the Spence model to an equilibrium with no discrimination is available from the author. Also, see Riley (1975) for a critique of the robustness of Spence's conclusions about suboptimality.

Although I do not find the empirical counterparts to the models of statistical discrimination and signaling to be convincing in terms of the necessary empirical magnitudes of such variables as costs of information or in terms of behavioral patterns, what is considered convincing and realistic is a matter of judgment. Some readers may not view trial work periods as realistic. A rigid system of "tracking" newly hired workers, for example, could scuttle the strategy of trial work periods. Others may believe that government and union wage floors are pervasive and, in combination with the statistical model shown in part (b) of Figure 13.1, block the employment of minorities on a large scale. There is need for institutional knowledge and for judgments.

3.4. Institutional theories of discrimination

In his survey of the economics of racial discrimination, Marshall (1974) advocated an institutional theory of discrimination which, although presented as an alternative to neoclassical theories, could be viewed as a plea for more complementary attention to such factors as historical contexts, "pre-labor-market" discrimination against minorities, group bargaining, the psychological motives of the economic agents, monopoly elements, and a variety of societal factors Marshall classified as environmental. Some points are well taken, and few neoclassical economists would argue in principle against them. Others reflect certain misunderstandings. Neoclassical theory is not, for example, synonymous with perfect competition; monopolies, including labor unions and governments, are not ignored in neoclassical economics. Pre-labor-market discrimination is allowed for in Model (I), represented by group differences in X between the minority and majority workers. Many of the societal factors Marshall mentions (1974, p. 868), such as health, education, and business conditions, have all received considerable attention in the neoclassical literature.

The institutional approach sometimes cuts across several disciplines. One example is the reference to psychology and theories of adaptive behavior. Piore (1970) argues that the initial placement of disadvantaged workers into low-wage, low-status jobs creates attitudes and habits that perpetuate their low status. Arrow (1973) suggested a related model in which the psychological theory of cognitive dissonance rationalizes market exchanges that result in a suboptimal equilibrium. In essence, expectations are formed by employers about the inferiority of the group discriminated against, and the latter internalize these expectations and take actions – in particular, underinvest in human capital – which confirm those expectations. An objection to both versions of this pattern of self-injurious behavior is that the predicted behavior is obviously counter to the best interests of two key actors – the group discriminated against, whose members want to overturn the expectations, and employers, who ought to prefer to

augment the supply of labor by encouraging more investment in human capital and positive attitudes towards investment and work.

Myrdal's (1944) classic work on discrimination included a similar model of feedback effects, in which economic, attitudinal, and health variables interact dynamically. An interventionist shock to any one of the variables sets in motion an upward or downward spiral of all the variables. Lundahl and Wadensjö (1984, pp. 16–18, 53) discuss Myrdal's model, including its similarity to Piore's, and point to its vulnerability to the criticism of instability. A corollary objection, similar to the one made in the preceding paragraph, is that the model's predicted consequences from a favorable shock are so obviously beneficial to the group discriminated against and to employers that it is difficult to see why the upward spiral would not quickly be initiated by group intervention. These criticisms apply, however, to the particular mathematical formulation of the model and not to the reasonable view that economic outcomes are determined by multiple causes, some of which are noneconomic, and that feedback relationships are part of reality.

If institutionalism refers to historical case studies, to details of the process by which equilibrium states (or tendencies) are reached, and to the interactions among organized and individual agents, then the approach – while not a theory, in my judgment – is always useful and sometimes indispensable. In the statistical studies that are discussed in some detail in the next section, there are often contexts in which various strata or segments of the full population are studied. The question arises: How were the individuals selected into these strata, and does the selection process either reflect discrimination or affect the interpretation we give to the analysis? For example, in studies of the effect of unions on the wage differential of black and white workers, institutional knowledge about the selection process into unions and how the process differs by race is necessary to interpret correctly the statistical estimations. Neoclassical economists are aware of the need for this information and, in one form or another, pay attention to the selection process [see Ashenfelter (1972), Becker (1959), Kessel (1958), Lewis (1959)], but they seldom have an absolute advantage in the institutional aspects of the problem. The legal and historical studies such as those of Gould (1977), Hill (1977), Marshall (1965), and Northrup (1944) are also useful.

Earlier, the institutional study by Ross (1948) was quoted to illustrate the harm done to black railway workers by the interactions of government wage-fixing (blocking the forces of wage competition), employer monopoly, Southern community prejudice, and an all-white labor union. Ross provides more institutional detail about this episode of discrimination that is worth retelling to remind those of us who work with austere models and simplified statistical specifications just how complex is the reality we are trying to capture. Ross recounts the advances made by black railway workers into the higher-paying jobs of firemen on the Southern railway carriers during World War I. Later, during two depression periods, 1921 and again in 1931, the white workers' grasp for these jobs reached

an intensity that took on an all too typically American climax. I quote Ross:

The depression of 1921 put many Negro and white workers on the street. There was violent competition to keep or grab places on any pay rolls. In 1921 there began a series of shootings from ambush at Negro firemen on Southern trains. Five were killed and eight wounded.... [In] the depression year of 1931...a Negro fireman, Clive Sims, was wounded on duty by a shot fired out of the dark beyond the track, the first of fourteen such attacks which stretched out over the next twelve months. This was not a racial outbreak in hot blood. It was a cold calculated effort to create vacancies for white firemen in the surest way possible, death, and, by stretching out the period of uncertainty and horror, to frighten away the others (pp. 119–120).

There are, as noted earlier, many theories or models that result in discriminatory outcomes. The challenge is to determine their quantitative importance. The instrument of terror, such as described above, no longer plays an important role in labor market discrimination. But even when this weapon is replaced with the milder instruments of racial and sexual harassments, we may find that the organized, sometimes conspiratorial, activities of majority workers and employers operate with a different set of rules than those we specify in our conventional economic models.

4. Empirical research on labor market discrimination

Aside from descriptive statistics, empirical research on economic discrimination may be divided into (1) tests of hypotheses suggested by the theories, such as the proposition that wage discrimination is less in competitive industries, and (2) estimation of the amount and determinants of discrimination; for example, estimating the effect of race on wages (the coefficient A) in a cross-section version of Model (I) (with productivity characteristics held constant), or estimating the change in the relative wages of minority workers over time.

4.1. Testing hypotheses suggested by theoretical models of discrimination

The hypothesis about labor market discrimination that has received the most attention is that discrimination is greater in monopolistic industries. An early empirical test is presented in Becker (1957, rev. ed. 1971, pp. 47–50). Many studies have followed.¹³ I do not review this hypothesis and these studies mainly

¹³ The following citations refer to studies, like Becker's, in which the proportion of minority-group employees in the firm, industry, or market is related to some measure of concentration (or degree of competitiveness): Comanor (1973), Oster (1975), and Luksetich (1979).

because I am uncomfortable with two links that connect the theory and the empirical evidence. First, I question whether product monopoly implies monopsony power in the labor market; the relevant labor market is usually a local area, and we have no assurance that monopsony power is highly correlated with the commonly used measures of monopoly, such as concentration ratios. This criticism, which applies to many of the previous studies, has been recently developed by Ashenfelter and Hannan (1986). Second, the desired theoretical measure of discrimination is the difference in minority/majority wages for equally productive workers, but most of the studies have used minority/majority employment differences (or ratios). While there is certainly interest in such employment ratios and associated measures of segregation as indicators of discrimination in the labor market, wage discrimination is not necessarily linked to segregation.

Aside from the studies of monopolies and discrimination, hypothesis testing has been, as Masters (1975, p. 19) noted, "surprisingly limited", and this type of study has produced few, if any, firm conclusions. In part this is because the theories often yield ambiguous predictions. For example, discrimination may be predicted to exist in the short run but not in the long run, but there may be no basis for determining the time required for the transition. Also, the theories suggest many economic influences, and the hypothesis test usually concentrates on one influence in isolation. The disappointing yield of most hypothesis testing may be conveyed by an examination of four studies.

(1) In his book, Reich (1981) criticized neoclassical theories of discrimination, provided tests of neoclassical hypotheses, and developed an alternative theory of discrimination that emphasized the role of class conflict between workers and capitalists. I focus solely on his test of Becker's model of employer discrimination in a competitive economy (pp. 109–163), which also appeared previously [Reich (1971)] and was discussed by Masters (1975, pp. 19–21). Reich claimed that Becker's model predicted a negative relation between (i) profits, which might more accurately be identified as the employers' return on their capital and their entrepreneurial skills, and (ii) the degree of discrimination, which is measured by and is inversely related to the ratio of blacks' wages to whites' wages, W_b/W_w , for equally productive black and white workers. An examination of Reich's analysis serves to illustrate several difficulties, listed as (a)–(c) below, in testing hypotheses.

(a) *The problem of ambiguity of theoretical predictions when, as shown in Section 3, there are many plausible outcomes, even within the neoclassical paradigm that Becker employed.* Reich claims that Becker's theory predicts that "white capitalists lose and white labor gains from racial discrimination" (1981, p. 111). This translates into a positive relation between profits and W_b/W_w . To see how this might occur, assume that white and black workers are equally productive, that their labor is inelastically supplied, that all employers have the same tastes for discrimination, and that employers' preferences for white workers lead to the

ratio W_b/W_w being less than 1. Now assume that the tastes of employers change to a stronger preference for whites. This leads to a higher wage for white workers and lower money profits for employers. The decline in profits is offset by a higher psychic income to employers from their enhanced preference for white workers, thus maintaining the total returns on their capital and entrepreneurial skills.

We here encounter a distinction, not emphasized earlier, between whether the employers' preferences are pro-white or anti-black. Had the hypothesized example assumed a change in preferences by employers toward greater distaste for black workers, then W_b/W_w would still decline, but in this case W_b would fall and money profits rise – the latter offsetting a decline in the psychic income of employers. A focus on the wage ratio leaves us with an ambiguous interpretation.

There are other sources of ambiguity. The observed variables are profits and wages, and these are predicted to change in response to an unobserved change in employers' tastes. However, the observed variables may change for other reasons, with a different application or interpretation of Becker's model. Assume now that there is variation in employers' tastes for discrimination, but that the distribution of employers' tastes does not change from one period to another. If the ratio of black workers to white workers increases, Becker (1971, pp. 43–45 and 97) predicts a fall in W_b/W_w , because the employers with stronger tastes against blacks can only be induced to hire the increased number of blacks by a decline in W_b . In this case, money profits rise, offset again by a fall in the psychic income of the new employers who are hiring blacks. Thus, the predicted short-run result is a *negative* relation between profits and W_b/W_w – opposite of the implication Reich draws from Becker's theory.

(b) *The problem of ambiguity because the predictions depend on the length of the time period to which they apply and because the theory offers no guidance on the time required for certain forces to take effect.* Reich's test of the relation between profits and W_b/W_w is based on a 1960 cross-section of 48 standard metropolitan statistical areas (SMSAs). Each SMSA is designated as a separate labor market. The hypothesis Reich is testing is one that assumes that employers' tastes vary across markets and that their tastes cause the variation in W_b/W_w . In 7 of his 43 reported regressions Reich (1981, pp. 135–155) controlled statistically for the ratio of the black population to the white population in the market, so this source of variation in W_b/W_w was, in principle, neutralized in these 7 regressions. In the other 36 regressions one could argue that more black workers lower W_b/W_w and increase profits and that this negative relation is consistent with Becker's model for reasons discussed above. The simple correlation between W_b/W_w and the percentage nonwhite in the SMSA is -0.71 in Reich's sample (1981, p. 149).

In the seven regressions in which the percentage nonwhite is controlled, Reich finds a negative relation between profits and W_b/W_w , but whether this is inconsistent with Becker's model depends, as we have seen, on whether one assumes variation in pro-white or in anti-black tastes among employers. Another

point is that in a cross-section any nonzero relation between profits and W_b/W_w may be viewed as a temporary disequilibrium, if the factors of production are mobile across SMSAs. Equally productive black (or white) workers would not remain in a market where they were underpaid relative to the wages available in other markets. Even though pervasive tastes against blacks by employers could lead to $W_b/W_w < 1$, the ratio should tend toward equality across markets if there is worker mobility. Alternatively, capital flows across markets will tend to equalize profit rates. If the profit variation is due to variation in employers' tastes, thereby allowing for the compensating variation in psychic income among employers, employers with the strongest tastes against blacks (or for whites) would tend to move to markets where blacks are relatively less numerous. Repeating the observation of Alchian and Kessel: "Free choice of economic activities implies a distribution of resources that would minimize the cost of satisfying tastes for discrimination" (1962, p. 161).

Neoclassical theories do not, however, tell us how long the equilibrating process will take, so tests involving SMSA data at a point in time could be thought of as either testing the competitive model or as testing the time of transition to equilibrium. Alternatively, a defender of a "sluggish" competitive model could test for the predicted equilibrating process by using SMSA data for two or more points in time.

(c) *The problem of matching the desired theoretical variables with the available empirical variables.* The hypothesis about the relation between profits and W_b/W_w for equally productive workers are actually tested by Reich by a regression between (i) a variety of measures of income inequality, such as the percentage share of all white incomes received by the top 1 percent of white families, S_1 , or the Gini coefficient of white family incomes, G , and (ii) the ratio of black to white family income, Y_b/Y_w . The Gini coefficient is a commonly used measure of overall income inequality, which includes the earnings of white workers. Becker's theory of employer discrimination made no prediction about the effect of W_b/W_w on the inequality of white workers' earnings. Nor is it obvious that S_1 is a good measure of profits, because the incomes received by the richest 1 percent of families will include rents, interest payments, wage and salary earnings, and income from inherited wealth as well as current profits from businesses employing workers.¹⁴

The theoretical variable, W_b/W_w , may diverge from Y_b/Y_w , and Reich provided no control for the relative productivities of black and white workers by such conventional measures as the ratios of mean educational attainments, mean years of experience, and so on. Reich's control variables were measures of the

¹⁴It should be noted that Reich expressed interest in the relation between discrimination and overall white inequality, so my discussion is restricted to Reich's use of these inequality measures to test Becker's model.

overall occupational and industrial structure, the median family income of whites (Y_w), the percentage of the SMSA population that is black (although in only one regression were both this percentage and Y_w included), and a few others. Generally, Reich found a statistically significant negative relation between Y_b/Y_w and his profit proxies, G or S_1 , which he interpreted as a refutation of Becker's model of a competitive economy and discrimination based on employers' tastes. In the light of the difficulties associated with items (a)–(c) above, I doubt that Becker's model was or can be well tested with such data.

(2) While Reich attempted to test for a relation between W_b/W_w and profits, sometimes controlling for the ratio of black workers to white workers, N_b/N_w , Landes (1968) and Flanagan (1973) drew upon Becker's theories to test for a negative relation between W_b/W_w and N_b/N_w . The justification from Becker's theory is as follows. Assume a distribution of employers' tastes for discrimination that is heterogeneous within a market and identical across markets. As we have seen, a larger N_b/N_w leads to a smaller W_b/W_w because the larger is N_b/N_w , the more are employers with stronger prejudices against blacks induced to hire black workers. The greater discrimination of these employers is manifest in a lower W_b/W_w , at least during the short run.

We have noted that mobility by black workers will tend to attenuate the negative relation between W_b/W_w and N_b/N_w , by tending to equate the ratios across markets. Also, there are institutional reasons for doubting the assumption of an identical distribution of tastes by employers across markets. Historically and in 1960, discrimination against blacks was most severe in the South, the region with the largest N_b/N_w . The legacy of slavery in the South was causal to both the discrimination and the residential location of blacks.

Scholars in other disciplines have debated how prejudice is related to N_b/N_w within a region. Perhaps prejudice is greater when N_b/N_w is greater because whites feel threatened by a larger ratio. On the other hand, perhaps the level of prejudice decreases as N_b/N_w rises because contact and familiarity erode unfavorable stereotypes and misunderstanding. In either case the level of tastes may change over time as experience with threats or with familiarity evolves. Thus, the basis for testing a version of Becker's theory that depends on identical distributions of tastes across markets appears questionable, although the empirical results of such tests are interesting on their own.

Landes (1968) found a negative correlation between W_b/W_w and N_b/N_w across all states, but the correlation was essentially zero within both the South and the non-South regions. However, this finding was secondary to Landes's main interest in the effects of antidiscrimination laws on W_b/W_w , so I examine the article by Flanagan (1973), whose main interest was to test the hypothesized negative relation between an occupation-specific W_b/W_w and an occupation-specific N_b/N_w . He used aggregated state data from the 1960 census for men in seven, and for women in five, broadly defined (one-digit) occupations. Other

variables in the regressions were the black-to-white ratios of four variables – weeks worked, educational attainment, age composition, and median family income – and two nonratio variables – a dummy variable for the South and the percentage of the population that was foreign-born. No systematic relation between W_b/W_w and N_b/N_w was found. This may be evidence against Becker's theory, or it may be evidence against Flanagan's maintained assumption that the distributions of tastes of employers are identical across states, or it may be that a simultaneous relation between wages (prices) and the quantities of occupational skills prevents the identification of an effect of the quantity ratios on the wage ratios.¹⁵

(3) A study by Chiswick (1973) is unusual for its focus on Becker's model of workers', rather than employers', discrimination and on wage inequality among whites – a topic not treated by Becker. Essentially, Chiswick tests the hypothesis that a measure of the variance of white male incomes in a state is positively related to the percentage nonwhite in the state.

Chiswick begins with Becker's definition of worker discrimination: a wage, W , is paid to (demanded by) a white worker who works with white workers, and $(W + d)$ is paid to (demanded by) a white worker who works with black workers. As we have seen in Section 3, segregation could prevent the long-run maintenance of wage discrimination against blacks, but Chiswick argues that inequality of wages is likely to persist if some white workers have skills complementary to the skills of black workers. Chiswick offers the example of "foremen" and "laborers", presumably where whites are both foremen and laborers and blacks are only laborers (p. 1332).¹⁶ Chiswick apparently rules out a segregated equilibrium in which there are some firms that hire only unskilled workers, who would be either all white or all black, and other firms that hire workers of both skills, who would be all white.¹⁷

Chiswick defines a dummy variable, X , as 1 if a white worker "works with nonwhites and ... zero if he does not" (p. 1333), and expresses the dual wage

¹⁵ Flanagan notes the potential simultaneity problem and refers in a footnote to his consideration of it. However, not enough information is provided to determine if the simultaneity problem is adequately handled.

¹⁶ The page numbers in parentheses in the text refer to Chiswick's article. Chiswick does not discuss the skill distribution of blacks or the possibility that blacks acquire skills. Note that if blacks acquire complementary skills, segregation could again eliminate racial discrimination in wages. Chiswick mentioned two other sources of integration in the work force besides complementarities in skills – unions and fair employment laws. The operations of these sources are not explained, except to note that they interfere with competitive market forces (p. 1332). Moreover, unions and fair employment laws are not mentioned again and play no role in Chiswick's empirical tests.

¹⁷ Firms employing all unskilled workers will pay equal wages to black and white workers. Firms employing any white skilled workers will hire only white unskilled workers to keep their costs at a minimum, so an equilibrium requires that all firms hiring both skills to hire only white workers. All unskilled workers, white or black, would receive the same wage. But this scenario merely reflects the segregation equilibrium that Chiswick has ruled out. Thus, we need to assume, as Chiswick implicitly does, that all firms require both skills.

structure for whites as $W^* = W(1 + d X)$, where W^* is the observed wage and W is the wage paid to the white worker who works only with whites. (A skill index, using a subscript for the j th skill, is omitted, and my symbols differ from Chiswick's.) The mean, \bar{X} , "is the proportion of the white labor force that works in an 'integrated' situation" (p. 1333), and Chiswick represents this by the percentage of nonwhites in the population, $p = (100)[N_b/(N_b + N_w)]$ (pp. 1334–1335).

The relationship between \bar{X} and p may be justified by assuming that unskilled workers have tastes for discrimination, so competitive forces should lead to their segregation by race.¹⁸ There would be no wage inequality among white unskilled workers (the laborers) within a market (or, for that matter, between markets – where a market is a state in Chiswick's formulation), at least as regards the effects of workers' tastes for discrimination. White skilled workers (foremen) would earn more if they worked in a firm with all-black unskilled workers than if they worked in an all-white firm, and labor costs would be equalized across firms by paying lower wages to black unskilled workers.

In this model and with the expectation that there are more firms with segregated unskilled workers in a state with a larger proportion of blacks, the mean wage of skilled workers should be positively correlated with p . This correlation identifies a direct test of Chiswick's model. A second direct test is the segregation of unskilled workers. I refer to these as direct tests because they involve cross-state comparisons of "first-order" effects on means and proportions rather than comparisons of "second-order" effects on within-state measures of inequality.

As noted in Section 3, the Becker-type models in which the skilled white workers have tastes for discrimination are similar to models with discriminating employers. Both agents are complementary to black labor. A long-run competitive equilibrium with discriminatory wage differentials paid to the skilled workers, like the long-run equilibrium with differential profits among employers, depends in both cases on homogeneity in the tastes of the discriminators. Or, expressed more cautiously, the tendency for discrimination to wither away depends on the existence of some nondiscriminating skilled workers (or employers) and on whether they can expand production to take advantage of their cost advantage.

Chiswick's empirical work focused on the variance of the logarithm of income for men aged 25 to 64, using midpoints of nine income classes, with an approximation for the mean of the highest, open-ended income class. This

¹⁸If the unskilled workers did not discriminate against each other, competitive forces would tend to make the proportion of black and white unskilled workers equal. Otherwise, either firms with more black workers would be at a competitive disadvantage – having to pay more to their skilled white workers – or blacks in firms with a larger proportion of blacks would be earning less than their counterparts in firms with a smaller proportion of blacks. See Arrow (1973, pp. 10–13) for a discussion of this case.

variable was regressed on p along with controls for several market sources of inequality in the form of variables involving the age, schooling, and weeks-worked distributions in the state and a variable defined as the rate of return on schooling in the state, which Chiswick had calculated in his previous research. Chiswick assumed that tastes for discrimination and p were uncorrelated. To make this assumption plausible, he separated the 17 Southern states from the non-Southern states. Chiswick found that white inequality was positively related to p , within both the South and non-South regions.

The causal inference seems shaky, but interpreting empirical tests that are indirect is always a matter of judgment. Here, p is an indirect measure of either the intensity of skilled workers' tastes against unskilled black workers or of the proportion of white skilled workers who receive higher wages by working with blacks, and the variance of income is an indirect measure of the skilled workers' wage inequality (since there should not be inequality among the white unskilled workers' wages). The regression for the South had only 8 degrees of freedom. In the 31 non-Southern states, there were only 13 where blacks were more than 3 percent of the population in 1960 [U.S. Bureau of the Census (1980, p. 36)]. The highest percentages, 8.0–10.0, were in the industrialized states: Illinois, Michigan, Missouri, New Jersey, New York, and Ohio. The lowest percentages, 0.1–0.9, were in relatively nonindustrialized states: Iowa, Idaho, Maine, Minnesota, Montana, New Hampshire, North Dakota, South Dakota, Utah, Vermont, and Wyoming. Thus, outside the South blacks were generally such a small proportion that it is difficult to see how they could have had much effect on white income inequality. Where they were a modest proportion, it was in states that tended to be more industrialized and densely populated.

Aside from how one might interpret Chiswick's regressions showing a positive relation between p and the variance of white incomes, I find them unconvincing as a test of Becker's model in the absence of direct information on how workers' tastes for discrimination affect (a) the segregation of workers and (b) the wages of white skilled workers who do and do not work with black unskilled workers. On this latter issue, Blau (1977, pp. 58–73) reports that in her study of labor market discrimination among several white-collar occupations, men who worked in integrated firms (with both men and women) received lower wages than men who worked in all-male firms, and she interpreted this as evidence against the hypothesis that workers' discriminatory tastes were causal to wage differentials.¹⁹ A problem with these tests, however, is the necessary assumption that the integration measure (say, the proportion of blacks or women in a firm) is uncorrelated with the average skill level of the white or male workers whose wage is the dependent variable.

¹⁹In private correspondence, Chiswick cites an unpublished study by James Ragan that also uses data for individual firms and finds higher wages for whites who work in integrated firms. This finding was interpreted as supporting the Becker-type model of worker discrimination.

(4) The final example of hypothesis testing is Ashenfelter's (1972) analysis of the effect of unions on the white-black and male-female wage differences. The model of discrimination under competitive conditions, which appears fragile and difficult to test in the previous examples, is replaced here by the more robust theory of union gains and a somewhat tentative theory of racial and gender selection into unions. Specifically, it seems reasonable to assume that union-based noncompetitive wage differences across racial and gender groups can be sustained. The effect of union status on a worker's wage is estimated by a Model (I) regression function, which is applied separately to the four race-gender groups. Each of the four union effects (coefficients) is multiplied by the percentage unionized of each race-gender group to show the difference in wages across the groups that is attributable to unionism.

A numerical example is helpful. Assume that the effect of unions is to increase the wages of unionized black men by 10 percent relative to nonunion black men, while the corresponding effect for white men is 5 percent. Assume also that the proportion unionized is 20 percent for both racial groups. A first approximation to the union effect on W_b/W_w is obtained by assuming that the wages of nonunion workers, W^n , are equal to what the wages would be in the absence of unions. Let the wage ratio for nonunion workers, black-to-white, be 70/100. This can be compared to an estimated ratio for all workers, holding constant available productivity characteristics. This estimated ratio is calculated as a ratio of weighted averages of the wages of union and nonunion workers, using the percentage union, $U = 0.2$, and the percentage nonunion ($= 0.8$) as weights. Thus the estimated ratio for all workers, holding constant their productivity, is

$$\frac{\hat{W}_b}{\hat{W}_w} = \frac{(1-U)\hat{W}_b^n + U\hat{W}_b^u}{(1-U)\hat{W}_w^n + U\hat{W}_w^u} = \frac{0.8(70) + 0.2(77)}{0.8(100) + 0.2(105)} = \frac{71.4}{101.0} = 0.707.$$

We see in this example that unions increase the overall wage ratio by 0.007, or by 1 percent, relative to what it would be in the absence of unions.

Clearly, the overall impact of unions on the majority-minority differential by these calculations depends on the percentage of each group that is unionized and the wage effect of unionism for each group. If the union effects for both racial groups are 10 percent and the proportion unionized is 30 percent for blacks and 20 percent for whites, the same impact of unions on the black-white wage ratio would be obtained.

Calculations like these were carried out by Ashenfelter, who first obtained estimates for U , \hat{W}^n , and \hat{W}^u for the four demographic groups. He added a refinement by computing estimates of \hat{W}^n and \hat{W}^u for major (one-digit) occupational groups and then summing these with weights for union and nonunion status that involve the proportion of the wage bill (total wages) received by each union-and-occupational group. Thus, instead of weighting the W 's by U ,

Ashenfelter used U^* , the proportion that union wages are of the total wage bill earned by whites (or blacks). The U^* values are larger than the U values, especially for blacks. The low percentage unionized of both blacks and whites in the higher-paying white-collar occupations carries a low weight for blacks relative to whites because relatively few blacks are in these occupations. Thus, although 23 percent of black workers in Ashenfelter's principal sample are union members, about 34 percent of the black wage bill is from black unionized workers. The comparable figures for whites are 23 percent and 31 percent.²⁰

Using U^* , Ashenfelter concluded that "the ratio of black to white male wages may have been some 3.4 percent higher in 1967 than it would have been in the absence of all unionism" (p. 463). The ratio of female to male wages was estimated to be 1.9 percent lower than it would have been in the absence of unions (p. 453, n. 33). The 3.4 percent gain to blacks reflects a differential effect of unions in favor of blacks by about 11 percentage points—a 21 percent effect for black men and a 10 percent effect for white men (p. 450). An illustrative weighted average for men is

$$\frac{\hat{W}_b}{\hat{W}_w} = \frac{0.66(70) + 0.34(84.7)}{0.69(100) + 0.31(110.0)} = \frac{75}{103.1} = 0.727,$$

which is 3.9 percent larger than the estimated wage ratio in the absence of unions, 0.7. Using the unrefined union weights, $U = 23$ percent for both blacks and whites, the weighted ratio would be 0.717, which is a little over 2 percent larger than 0.7.

These findings are evidence against the hypothesis that unionism in the United States, as measured during the 1960s, is responsible for the discriminatory wage differential in favor of whites or, with weaker evidence, in favor of men. The data on union membership by demographic groups are not controversial, and Ashenfelter provides alternative estimates of the effects of unions on wages, based on his own analysis of other data sources and on the existing literature. Overall, these checks were supportive. Ashenfelter reminds the reader that his evidence does not say that unions are nondiscriminatory; rather that they are shown to be no more discriminatory, or even less regarding blacks, than the economy as a whole.

The validity of Ashenfelter's estimates of union effects depends on two key assumptions. The first is that the estimates of union effects on union workers are either unbiased or that they are biased equally for majority and minority groups (hereafter, white and black men). The general issue concerning a bias is that

²⁰ These percentages are calculated using Tables 6 and 7 in Ashenfelter (1972), although I adjusted the weights in Table 7 for whites to make them sum to 1. Apparently there is an error in Table 7 for white workers, because the proportions sum to 1.072 instead of 1.00. I reduced each occupation's proportion in the table by 0.92 ($= 1.00/1.072$). In my calculation of U^* I assume that the percentage unionized for private household workers and farm workers is zero for both color groups.

union status may be correlated with unmeasured productivity variables, leading to a misestimate of the true effect of unions. As stated, the bias could be positive or negative, depending on whether union workers were, holding constant the control variables in the model, less productive (owing to, say, nepotism or perhaps because unambitious workers are more attracted to unionism) or more productive (due, say, to the commitment of union workers to their trades or because employers will select high-quality workers when faced with union-imposed above-competitive wages and because high-quality workers will seek these positions). To sharpen my argument and shorten the discussion, let me assume that the net bias in the union effect is positive, and the coefficients of union status on wages, 10 percent for whites and 21 percent for blacks, are both too high. Clearly, the issue for Ashenfelter's measure of the union impact on W_b/W_w is whether the bias is larger—really, much larger—for blacks than whites.

I now argue that the bias is larger for black men and against Ashenfelter's assumption that wages of nonunion workers represent what the wages would be in the absence of unions. Assume that the jobs in the union sector are medium-paying jobs in the crafts and operative occupations for both white and black men, whereas jobs in the nonunion sector are predominantly high-paying professional, technical, managerial, and sales jobs for white men, but predominantly low-paying laborer and unskilled service jobs for black men. Skill levels of the jobs are assumed to be correlated with the skill abilities of workers, both innate and acquired. Assume that these contrasting alternatives to whites and blacks regarding nonunion jobs are entirely attributable to "pre-labor-market discrimination", which is to say that they are reflections of differential family socioeconomic backgrounds, quality and quantity of schooling, and wealth constraints on the long-term investments required for the high-paying jobs. Assume further that the distributions of innate ability (intelligence, "ambition", and so on) are identical for whites and blacks. Given these assumptions, it is reasonable to believe that if unions were nondiscriminatory then black males would be more represented in the union jobs because black workers of above-average ability are constrained from entering the highest-paying jobs but not (by assumption) from crafts and operative jobs, and they will therefore gravitate toward the better-paying crafts and operative jobs. The presence of unions restricts numbers of both blacks and whites, but the restrictions are more binding on blacks, since the excluded higher-ability white workers will have the highest occupations open to them. Thus, not only should the unionized percentages be higher for blacks in a nondiscriminatory labor market, but the effect of union status on wages will tend to be more upward-biased for blacks. The latter bias stems from the presumption that the omitted innate ability is, on average, higher for black union members than for white union members. Ashenfelter's model assumes that the occupational distribution of blacks and whites is given, and the foregoing argument suggests that it is affected by unionism.

The arguments above are admittedly speculative. An upward bias in the estimated union effect, however, has the theoretical justification that employers should respond to union-imposed high wages by upgrading their hiring and retention standards. Generally, unionized employers do have control over hiring, and they have some control over retention, at least through some probationary period before union-imposed seniority protection commences. On the other hand, arguments in favor of Ashenfelter's conclusion are the following. (a) The above scenario denied any role to labor market discrimination for the disproportionately low representation of black men in the white-collar occupations, and this denial is hard to accept. (b) For a reduction in unionism to lead to relatively more occupational upgrading among blacks than whites among the blue-collar occupations, one must assume that the general sources of labor market discrimination would not maintain the existing distribution. (c) Ashenfelter's estimated union effects on wages would have to be drastically changed to reverse his conclusion of a beneficial wage-effect for blacks among unionized workers. Recall that his union effect for blacks (21 percent) is twice that for whites (10 percent).

Finally, Ashenfelter's rejection of the hypothesis that W_b/W_w would be higher without unions is strengthened by his institutional and historical discussion about union race policies. For historical reasons, unionism is more widespread among the blue-collar occupations, and blacks are more likely to be competing for jobs requiring less skill. Thus, Ashenfelter argues that because unions of lesser-skilled workers will have more blacks in their jurisdiction, competitive forces will tend to force the unions to include blacks. Among blue-collar occupations, therefore, whites will be over-represented in the unionized skilled jobs and underrepresented in the unionized lesser-skilled jobs, relative to blacks. An overall tendency for equality in the incidence of union membership among whites and blacks is, therefore, plausible.

The remaining parameter of interest, the union-effect differential, is, however, puzzling on theoretical grounds. Ashenfelter's arguments (p. 447) about the potential power of the skilled trades to be more restrictive in controlling the supply of labor should be supported by larger union effects (rents) for the more skilled groups. This result would indeed be consistent with the *a priori* Marshallian arguments, found in almost every labor economics textbook, in which skilled workers face a more inelastic demand curve and therefore have more "bargaining power". As noted above, this result is not found by Ashenfelter, nor by other recent analysts of union effects. [See the studies Ashenfelter cites in his Table 3, p. 446; Johnson (1975), and others.]²¹ Thus, the large union

²¹ To be more precise, my claim is that Ashenfelter and others have found an overall negative correlation between union effects and skill levels, even though the construction trades, airline pilots, and some other crafts have shown large union effects. Among white construction workers, incidentally, the union effects of laborers exceed those of skilled workers [see Ashenfelter (1972, Table 5, p. 450)].

effects for blacks, relative to whites, is consistent with the larger union effects for lesser-skilled blue-collar workers, but the latter union effect remains a puzzle.

4.2. *Estimating labor market discrimination*

4.2.1. *Methodological points*

Model (I) (in Section 2) is the basic model used to estimate labor market discrimination. Its widespread use along with several conventions that are customarily adopted permits a succinct summary of results, shown in Tables 13.6 and 13.7 in the next section. Unfortunately, the results are so varied that they reveal as much about our ignorance as about our knowledge of the degree of labor market discrimination against blacks and women. This variability is not really surprising in light of the theoretical vagueness that underlies most of the empirical specifications.

An inherent ambiguity, mentioned earlier (Section 2), stems from the absence of agreement on what productivity traits—the X 's in Model (I)—are appropriately held constant. The criterion I suggested is that the variables held constant in Model (I) should *not* be determined by the process of discrimination under analysis. Applying the criterion requires a clear statement of the purposes of the estimations, but this is seldom provided. Perhaps the marketwide regression studies of wage discrimination are merely intended to provide a general social indicator of inequity in the economy, although this is ambiguous unless we know what counterfactual regime is being compared to the current regime. This counterfactual is usually only implicitly revealed by the set of X -variables that have been held constant, and there is seldom discussion of whether the X 's are affected by labor market discrimination.²² Predictions using the regression results are not often explored, and specific remedies or policies to deal with discrimination are seldom linked to the regression results. To clarify some of these issues, consider the following two applications of the criterion suggested above.

Case 1. Assume the analysis pertains to a given employer or firm, and that we ask whether white workers are paid more than black workers after holding constant the available productivity variables. Assume further that a panel of experts provides us with the worker characteristics that determine productivity in the firm. The productivity variables might include previous vocational training, tests of manual dexterity, age, years of schooling, and so on. To meet the above criterion, each variable should be exogenous to the employer; that is, the

²² Blinder (1973) is exceptional in his clear distinctions between the X 's that are assumed exogenous and those that are endogenous according to current theories of labor market behavior, specifically the theory of human capital.

characteristic should not be affected by the employer's behavior. If it were, it might reflect discrimination. Thus, a variable defined as "task-specific ability" that is measured by "supervisor's rating" would be suspect, and perhaps not admissible. Clearly, the presumptive identification of supervisors with management raises suspicions about the unbiasedness of supervisors' ratings. On the other hand, if we knew that supervisors were nondiscriminatory, their ratings would provide direct evidence of the workers' productivity, which is usually difficult to obtain and certainly preferable to the indirect evidence from such variables as age and education.

Case 2. Assume the analysis pertains to the entire labor market. We ask whether white workers are paid more than black workers after holding constant an admissible set of productivity variables that are not affected by the process of discrimination under analysis. Because the entire labor market is under analysis, however, variables like "previous training" almost surely reflect previous discrimination in the labor market, so they are not admissible.

There is no simple rule in marketwide studies for determining when a variable may be appropriately held constant. Among the variables mentioned in Case 1, age is clearly exogenous. Years of schooling are appropriately held constant if we believe that the decision to attain schooling does not reflect discrimination in the labor market. Perhaps the lower education among minorities reflects societal discrimination – not labor market discrimination but "pre-labor-market discrimination". Alternatively, perhaps blacks and women perceive that higher levels of schooling yield smaller earnings for them than for white men. If this were true, then these groups may have curtailed their schooling, in which case educational attainment would reflect labor market discrimination. Determining the productivity variables that are admissible is the first step in estimating Model (I). Accurate measures of the agreed-upon variables are also needed.

Let us turn now from the conceptual issues in estimation to the mechanics of the statistical methods. The regression specifications for Model (I) that produce the estimates of labor market discrimination in the recent research literature usually involve the following assumptions and procedures.

(i) Separate regression functions are estimated for majority (hereafter white, w) and minority (hereafter black, b) groups. The equations omit the subscripts denoting the observation. The explanatory X -variables measuring productivity traits, and the B -coefficient of each X are collectively represented by $\sum BX$.

$$\hat{W}_w = \sum B_w X_w \quad \text{and} \quad \bar{W}_w = \bar{\hat{W}}_w = \sum B_w \bar{X}_w, \quad (5)$$

$$\hat{W}_b = \sum B_b X_b \quad \text{and} \quad \bar{W}_b = \bar{\hat{W}}_b = \sum B_b \bar{X}_b. \quad (6)$$

The caret indicates predicted value, the mean of which, $\bar{\hat{W}}$, is identically equal to

the overall mean, \bar{W} . The intercept term in the equation is included in $\sum BX$ and may be associated with an element in the X -vector for which $X_b = X_w = 1$ for each observation.²³

(ii) Equations (5) and (6) are used to express eq. (7), which is a particular decomposition of the difference in mean wages obtained by adding the term $\sum B_w \bar{X}_b$ to both (5) and (6) and then subtracting (6) from (5):

$$\bar{W}_w - \bar{W}_b = \sum B_w (\bar{X}_w - \bar{X}_b) + \sum \bar{X}_b (B_w - B_b). \quad (7)$$

The firm term on the right-hand side of (7) evaluates the difference in mean values of the X 's at white "prices" (B_w 's), and the second term evaluates the racial price differences at the mean value of the black X 's. It turns out that, on average, $\bar{X}_w > \bar{X}_b$ and $B_w > B_b$; more precisely, that $\sum B_w \bar{X}_w > \sum B_b \bar{X}_b$.

(iii) The second term on the right-hand side of (7) is a conventional measure of labor market discrimination, with $B_w > B_b$ representing a higher price received by a white worker than by a black worker for the (assumed) same productivity characteristic. The first term on the right-hand side of (7) involves the racial differences in X 's and does not have a clear interpretation. It may represent a source of a nondiscriminatory difference in wages, because only one price is used to evaluate different amounts of exogenous productivity characteristics. Or, it could measure the difference in wages attributable to pre-labor-market discrimination, which may explain why $\bar{X}_w > \bar{X}_b$. In any case, the conventional standard of nondiscrimination is achieved when $\hat{W}_b/\hat{W}_w = 1$, holding the X 's constant.

(iv) An important reservation about the decomposition in (7) is that it is not unique. Each difference, $B_w - B_b$, in the second term is evaluated as a product with \bar{X}_b , but the evaluation might have used \bar{X}_w or some average of \bar{X}_b and \bar{X}_w . Similarly, the use of B_w as a weight for the first term, $\bar{X}_w - \bar{X}_b$, is also arbitrary. A different decomposition of $\bar{W}_w - \bar{W}_b$ is obtained by adding the term $\sum B_b \bar{X}_w$ to both (5) and (6) and subtracting (6) from (5):

$$\bar{W}_w - \bar{W}_b = \sum B_b (\bar{X}_w - \bar{X}_b) + \sum \bar{X}_w (B_w - B_b). \quad (7')$$

The different standardizations shown by (7) and (7') reflect the familiar index-number problem encountered whenever heterogeneous collections of goods

²³Blinder (1973, pp. 438–439) separated the intercept terms from other B -coefficients and specified them as two components of discrimination. This procedure is not necessary or even helpful, because the value of the intercept term will depend on the arbitrary scaling of the X -variables. Consider, for example, the arbitrariness of defining a variable like region of residence into a set of dummy variables, where the intercept will represent the excluded region. Which region is to be excluded is arbitrary. See Jones (1983) for further discussion of the point.

(X 's) are summed with two sets of prices (B 's).²⁴ In the simplest case in which all prices are the same for both racial groups, the difference $\bar{W}_w - \bar{W}_b$ is simply equal to the first term on the right-hand side of (7), and the conceptual experiment of assigning equal X 's to both racial groups leaves only the difference in intercept terms, which measures a vertical difference in \hat{W} between two "parallel" linear functions. Such a difference in intercept terms is what was previously measured by the coefficients A or C on the dummy variables for group status in Models (I) and (II) in Section 2.

(v) I will rely on the following expressions for summarizing the various estimates of eqs. (5) and (6) reported in the literature:

- (i) $U_r = \bar{W}_b / \bar{W}_w$ = the unadjusted ratio,
- (ii) $A_r = \sum B_b \bar{X}_w / \sum B_w \bar{X}_w = \sum B_b \bar{X}_w / \bar{W}_w$,

which is an "adjusted ratio", obtained from either (7) or (7'). To arrive at (ii), simply set (or assume) all \bar{X}_b equal to \bar{X}_w to eliminate the first term of the decompositions in (7) or (7') and to reduce the right-hand side to its discrimination component. Then divide through by \bar{W}_w and simplify to express the following equation of ratios: $A_r = \bar{W}'_b / \bar{W}_w$, where \bar{W}'_b is the black mean wage conditional on the black \bar{X} 's being set equal to the white \bar{X} 's. $A_r = 1$ implies no discrimination. The amount by which the controls for X have closed the gap between unity and U_r is the sometimes-used statistic

$$(iii) \quad G = [(1 - A_r) / (1 - U_r)] 100,$$

called the percentage of the gap between U_r and 1 that is attributable to the difference in the X 's. Thus, $1 - G$ is the percentage of the gap that is attributable to labor market discrimination.

For simplicity, I will restrict my discussion of empirical results to the adjusted and unadjusted ratios, A_r and U_r . Even here, it is somewhat arbitrary to use A_r as defined by (ii), because we could have defined

$$(ii') \quad A'_r = \sum B_b \bar{X}_b / \sum B_w \bar{X}_b = \bar{W}_b / \sum B_w \bar{X}_b,$$

which, like A_r , holds the \bar{X} 's constant and attributes the remaining differences in black and white average wages to the B 's, but here the B 's are multiplied by \bar{X}_b

²⁴ Blinder (1973, p. 438, n. 3) suggested that the decomposition expressed by eq. (7) is preferred to that of (7') because he claimed that the decomposition using black prices (B_b 's) as weights for the difference in X 's leaves an interaction term as a residual, in contrast with the decomposition using white prices (B_w 's) as weights. This is incorrect. There is no difference in the two decomposition methods in this respect.

levels. Usually A_r is presented, because the conceptual experiment of raising \bar{X}_b to the levels of \bar{X}_w is more appealing and more policy-relevant than lowering \bar{X}_w to \bar{X}_b levels as is done with A'_r . Nevertheless, it is easy to construct examples in which the regression results give qualitatively different measures of wage discrimination on the basis of A_r and A'_r . One may equal unity and the other may exceed or fall short of unity. The quantity $A_r - U_r$ may be positive, showing that the X 's "explain" some of the gap (assuming $U_r < 1$), whereas $A'_r - U_r$ may be negative, showing that the gap is made even wider after controlling for the X 's and using this standardization.

At the risk of belaboring the obvious, I will use the constructs of A_r and A'_r to illustrate two points. One is the potential ambiguity of these ratios as measures of discrimination, and the second is that some institutional knowledge of the process by which discrimination occurs is necessary if the statistical measures are to tell us anything.

Assume eqs. (8) and (9) refer to males (subscript m) and females (subscript f) and that the only explanatory variable, X , in the wage function is the number of young children present in the household of the worker. The wage functions, evaluated at means, are

$$\bar{W}_m = B_{0m} + B_{1m}\bar{X}_m = 10 + 1(\bar{X}_m = 2) = 12, \quad (8)$$

$$\bar{W}_f = B_{0f} + B_{1f}\bar{X}_f = 9 - 1(\bar{X}_f = 1) = 8. \quad (9)$$

In eq. (8), I assume that all men are working, that they are in families with an average of two young children, and that the presence of young children has a positive effect on the wages earned by the men. (Perhaps additional dependents lead them to work harder.) Eq. (9) is assumed to express the wage equation for employed women. I assume that half the women are employed (and thus have a market wage), that women with fewer young children are more likely to be in the labor force, and that the presence of children is negatively related to the wages of women, which is discussed below.

Clearly, $A_r = \sum B_f \bar{X}_m / \bar{W}_m = 7/12 = 0.58$, and $A_r - U_r = 0.58 - 0.67 = -0.09$. Thus, the unadjusted ratio is higher than the adjusted ratio. The women's wage would be less than their current wage if they had the same values of X as men, so we may conclude that discrimination is even more severe than shown by the unadjusted wages.

On the other hand, $A'_r = \bar{W}_f / \sum B_m \bar{X}_f = 8/11 = 0.73$, and $A'_r - U_r = 0.73 - 0.67 = 0.06$. This shows that discrimination against women would be less if men had the same values of X as women. Since the regression method shows that discrimination is both worse and better than the unadjusted wage comparisons, what should we conclude? Or, consider the following specification:

$$\bar{W}_m = 6 + 1(\bar{X}_m = 2) = 8 \quad (10)$$

and

$$\bar{W}_f = 9 - 1(\bar{X}_f = 1) = 8. \quad (11)$$

A_r ($= 0.875$) shows discrimination against women, whereas A'_r ($= 1.14$) shows discrimination against men.

The statistical procedures cannot tell us the correct answer. Let us consider two hypothetical processes by which employers pay wages to examine how we might determine whether there is discrimination.

Case 1. Assume men and women are equally productive, but that employers discriminate against women with children, as shown in (9) or (11). Let us assume that children have no real effect on productivity of either men or women but that employers have a uniform preference for paying men with children more and women with children less. Both sets of equations, (8)–(9) and (10)–(11), show this discriminatory behavior, and constructing ratios and making decompositions do not add to our knowledge.

Equations (10)–(11) do not show a gender difference in average wages, but one could argue that the discrimination against women with children expressed in (11) is a cause of the low labor force participation of women. Assume that all women enter the labor market and that demand conditions remain the same. Then (11) becomes $\bar{W}_f = 9 - 1(\bar{X}_f = 2) = 7$, and U_f becomes $7/8 = 0.875$. The discrimination, expressed in (11) and revealed by a “potential” $U_f < 1$ for the full population, is a cause of the fact that only half the women are working. This is a reminder that the values of the X 's may reflect discrimination and are not always exogenous to the process under study.

Case 2. Assume that men and women have the same productive capacity, but women with children are less committed to market work than women with no children, and men with no children are less committed to their jobs than men with children. If these commitments reflect people's preferences about how they want to live, and if the presence of children is an accurate signal of this commitment, then there is no presumption of discrimination. The two sets of equations show the relation between productivity (commitment) and children and also show how the relation differs for men and women. Again, the ratios A'_r and A_r do not add anything useful to our knowledge.

The message is that the original data and statistical functions are mere description. Knowledge of the process by which wages are set and, perhaps, by which workers are selected into the market are necessary.

4.2.2. *A survey of selected estimated wage functions*

In contrast to the research that tests hypotheses, the studies presenting empirical estimates of labor market discrimination are numerous. Only about 20 of these

studies are selected for mention in this section, and they will be summarized in two tables. Methodological issues are emphasized to aid in understanding the strengths and weaknesses of the research and its theoretical and policy content. In addition, these empirical studies contain useful descriptive statistics.

Labor market discrimination, or wage discrimination, has been defined in this paper by using Model (I) to isolate the net effect of minority-group status on wages, holding constant the productivity characteristics of the workers. Two crucial questions invariably arise: (a) Do the variables measuring productivity reflect discrimination? (b) Do the variables measure productivity comprehensively, aside from factors that can be assumed to be random with respect to group status? If the answer to (a) is yes, we may presume that the estimate understates discrimination. If the answer to (b) is no, the estimate of discrimination may be biased up or down. To avoid prejudging these answers, I will use the term "wage gap" rather than "wage discrimination". The wage gap will be measured by the unadjusted ratio, U_r , and by the adjusted ratio, A_r (or, rarely, A'_r).

The estimated wage gaps are based on cross-section or time-series studies, which were the classifications used in the descriptive statistics presented in Section 2. Cross-section studies generally are interpreted as representing normal or equilibrium conditions. Trends over time may be inferred from successive cross-sections, allowing for changing compositional effects (like the age distribution) or specific period effects (such as the business cycle). Trends may be directly measured in a time series by introducing time as an independent variable and determining if its effect differs for the different groups, but time-series studies are hampered by the fewness of observations. Almost all the empirical work in the published literature uses cross-sectional data, and this section will be devoted to these, leaving the few time-series studies for the final section on policy analysis.

Another classification of the studies is by the type of minority and majority groups being compared, and I will continue to focus on black-white and women-men comparisons. Some estimates of wage gaps according to national origins and religions will be briefly mentioned.

I also concentrate on studies that intend to estimate the overall wage gap, rather than on studies that focus on the differential effects on wages of particular variables, like education, years of work experience, union status, or participation in some government program. Finally, I concentrate on studies that measure the wage gap for the entire labor force, or at least for large groups in the labor force. Only limited attention is given to the many studies of the wage gap within individual firms or within occupations.

4.2.2.1. Comparisons of the earnings gap between women and men. A summary of studies of the wage gap between women and men is shown in Table 13.6. The wage is, as discussed below, the most appropriate simple measure for examining

Table 13.6
Summary of studies of ratios of women's earnings to men's earnings, unadjusted and adjusted for various characteristics of workers and jobs.

Author and year of publication ^a	Data source and population studied ^b	Measure of earnings ^c	Statistical method and explanatory variables ^d	Women's earnings as a ratio of men's	
				Observed ^e $= U_r$	Adjusted ^f $= A_r$
Gwartney and Stroup (1973)	Census, age 25+ with positive incomes	\bar{y} , 1959 \bar{y} , 1969	T, R: 1, 2 (grouped data) R, S: 1, (2), 7, 23	0.33 0.32	0.39 0.40
Featherman and Hauser (1976)	OCG, married workers	y , 1961 1972	R, S: 1, (2), 7, 23	0.38	0.48*
Blinder (1973)	PSID, white working household heads and spouses, age 25+	w , 1969	R, S: 2, (3), 9, 12-14, 32, 34	0.54	0.54
Sawhill (1973)	CPS, wage and salary workers, age 14+	y , 1966	R: 1, 3, 10, 13	0.46	0.56
Gwartney and Stroup (1973)	Census, age 25+, full-time, year-round workers	\bar{y}_f , 1959	T, R: 1, 2, (10)	0.56	0.58
Suter and Miller (1973)	NLS, CPS wage-and-salary workers, age 30-44	y , 1966	R, S: 1, (2), 6, 10, 23	0.39	0.62*
Roos (1981)	GSS, white workers, age 25-64	y , 1974-1977	R, S: 1, 2, 10, 22, 23, 26, 29-31	0.46	0.63*
Fuchs (1971)	Census (1/1000 sample), nonfarm workers	w , 1959	R: 1, 2, 3, 8, 12, 25, 33	0.60	0.66
Treiman and Terrell (1975)	NLS: married workers, age 30-44: white nonwhite	y , 1966	R, S: 1, (2), 6, 7, 10, 17, 22	0.42 0.54	0.67 0.68
Cohen (1971)	Survey of working conditions, full-time nonprofessional wage-and-salary workers, age 22-64	y_f , 1969	R, S: 1, 2, 10, 11, 16, 24, 27, 28	0.55	0.69
Blinder (1973)	PSID: white, working household heads and spouses, age 25+	w , 1969	R, S: 1, 2, (3), 5, 9, 11-14, 21, 27, 32, 34	0.54	0.70*
Oaxaca (1973)	SEO, urban workers, age 16+ white nonwhite	w , 1967	R, S: 1(2), 3, 7-10, 12, 13	0.65 0.67	0.72 0.69
Sanborn (1964)	Census, wage and salary workers	\bar{y} , 1949	T: 1, 2, 3, 10, 18, 20	0.58	0.76*
Oaxaca (1973)	SEO, urban workers, age 16+ white nonwhite	w , 1967	R, S: 1, (2), 3, 7-10, 12, 13, 21, 25-27	0.65 0.67	0.78* 0.80*

Table 13.6 continued

Author and year of publication ^a	Data source and population studied ^b	Measure of earnings ^c	Statistical method and explanatory variables ^d	Women's earnings as a ratio of men's	
				Observed ^e = U_r	Adjusted ^f = A_r
Kohen and Roderick (1975)	NLS, full-time wage and salary workers, age 18-25	w_f , 1968-1969	R, S: 1,3,4, 7-9,13-15		
				white 0.76	0.78
Mincer and Polachek (1974)	NLS, SEO, white wage and salary workers, age 30-44	w , 1967	R, S: 1,(2),(3) 6,11	black 0.82	0.81
				married 0.66	0.80
Corcoran and Duncan (1979)	PSID, working household heads, age 18-64, white	w , 1975	R, S: 1,(2), (3),5,6,9, 11-13,16,17	single 0.86	0.87
					0.74
Sanborn (1964)	Census, wage and salary workers	\bar{y} , 1949	T: 1-3,6,10, 16,18-20,24	0.58	0.88*
Malkiel and Malkiel (1973)	Professional full-time employees in one company	annual salary 1966, 1969-1971	R, S: 1,6,8,16, publications, Ph.D., field	0.66*	0.77*
Astin and Bayer (1972)	Survey of college faculty	annual salary	R: rank, degree, field, research output, type of college	above + job level ^g 0.66*	0.86*
				0.78*	0.87*
Johnson and Stafford (1974)	Survey of Ph.D.s in college faculties	9-month salary	R: years since degree, field, sector, experienced (= 10 years)	0.85*	0.93*

Sources: D. J. Treiman and H. I. Hartmann, eds., *Women, Work, and Wages: Equal Pay for Jobs of Equal Value* (Washington, D.C.: National Academy Press, 1981), pp. 20-37, along with additional material and changes in the tables presented in Treiman and Hartmann.

^aFull citations are given in the references. The same study may appear more than once in the table.

^bSources for the individual studies use the following shorthand terms:

Census = the decennial Census of the United States.

CPS = the Current Population Survey of the U.S. Bureau of the Census

OCG = the CPS survey of Occupational Change in a Generation, 1962 and 1972

GSS = General Social Survey, 1975-1978

SEO = Survey of Economic Opportunity, 1966-1967

NLS = National Longitudinal Survey, 1967 and subsequent years (Ohio State University)

PSID = Panel Study of Income Dynamics, 1968 and subsequent years (University of Michigan)

Terms such as "aged 25+" refer to "workers aged 25 or older", and so on.

^cy = annual earnings (or income)

^d \bar{y} = a group's mean (or median) annual earnings (or income)

^ew = wage per hour

^fsubscript f = refers to full-time workers

^dT = tabular standardization

R = regression analysis

S = regression analysis using separate equations for men and women

Explanatory variables are listed by number at the end of these notes. The use of parentheses around a number indicates that this variable is implicitly held constant, either because of the sample selection or because another variable effectively controls for the variable in question. For example, if only whites are sampled, then race is being held constant.

^e U_f = the ratio of mean female earnings (or income or wage) to mean male earnings.

^f A_f = adjusted mean-earnings ratio, which is the ratio of the conditional mean earnings of women to the mean earnings of men. The conditional mean earnings of women is the earnings predicted for women if they had the same values of the explanatory variables as do men.

^gThe term "above +" means that the explanatory variables used are the same as those in the preceding list, plus whatever new variables are listed.

* The explanatory variables include a control for the occupation of the worker. Controlling for occupation is especially likely to raise the ratio of women's to men's earnings, for reasons discussed in the text.

Explanatory variables:

1. Education
2. Age
3. Race
4. Mental ability (intelligence)
5. Formal training
6. Actual labor market experience
7. Proxy for labor market experience
8. Marital status
9. Health
10. Hours of work (annual, weekly, full-time/part-time)
11. Tenure (length of service with current employer)
12. Size of city of residence
13. Region of residence
14. SES background (parental education, occupation, income, number of siblings, migration history, ethnicity, etc.)
15. Quality of schooling
16. Absenteeism record
17. Dual burden (number of children, limits on hours of location, plans to stop work for reasons other than training, etc.)
18. Urban/rural
19. Turnover
20. Occupation (census three-digit)
21. Occupation (census one-digit)
22. Occupational prestige
23. Occupational SEI (Duncan scale of a socioeconomic index)
24. Other occupational classification or scale
25. Class of worker (self-employed, government, or private wage and salary)
26. Industry
27. Union membership
28. Type of employer (government/private, sex segregated/integrated, size of work force)
29. Supervisory status
30. Percentage female in work group
31. Median income of male incumbents
32. Local labor market conditions
33. Length of trip of work
34. Veteran status
35. Migration status

labor market discrimination between men and women, but most of the studies use earnings or incomes. For brevity, I will refer to the earnings gap. The style and much of the content of the table are taken from the compilation of studies in Treiman and Hartmann (1981). The columns denote the authors of the studies, the data sources, the measure of the dependent variable, the statistical method (usually regression analysis) and the control variables used, the unadjusted ratio, U_r , and the adjusted ratio, A_r , which is the ratio of the average predicted earnings of women to the average earnings of men. For A_r , the earnings of women are usually predicted by a regression equation. When separate regressions for men and women are used, the earnings of women are predicted by assigning men's mean values for the predictor variables along with the regression coefficients from the women's equation.

The studies are listed in rough order of the size of A_r , which, although cautiously referred to as the wage gap in the sample, given the particular control variables used, is sometimes referred to as a measure of labor market discrimination. An asterisk next to the ratio indicates that some measure of occupational status was held constant. Among all the commonly used control variables, occupation is perhaps the one most "suspect" or "tainted" as being a reflection of labor market discrimination. It is an inappropriate control variable by the criterion I have proposed, although as noted above almost any variable that is subject to some choice by the individual worker and to some influence by the market may be suspect according to this criterion. Of course, occupation may have been advisedly included because the investigator wanted to measure the wage gap conditional on being in a given occupation. Thus, the asterisk is not an indicator of a defective study but rather of a study that does not measure marketwide discrimination as I have chosen to define it.

Reading down the rows of Table 13.6, we see that U_r and A_r have a similar ranking, and both range from 0.3 or 0.4 to around 0.8 or 0.9. The high figures usually refer to restricted samples. Much of the variation in these estimates may be explained in common-sense terms according to the following characteristics of the studies.

(1) *The use of earnings* (or in rare cases income) for persons in the labor force tends to give a lower ratio than the use of wage rates. The latter holds constant the unit of time for which earnings are measured. I prefer the hourly wage for gender comparisons because the amount of time spent at work will partly reflect voluntary choice. In contrast, an earnings comparison may be more useful for comparing white and black men, because working less than full-time by black men often reflects discrimination rather than voluntary choices.

(2) *Samples that represent the full population* generally show a smaller ratio. There is no necessary reason for this pattern; rather, the restricted samples happen to be for groups where the gap is narrower, such as for young age groups, for single women, or for certain occupations or industries. The wage gap is

narrow for young people, and it widens with age. This could mean that there is little discrimination by gender for young people and that the widened gap among older workers merely reflects the voluntary choices of women and men to specialize later on in housework or market work, respectively. Or it could mean that discrimination takes the form of providing women fewer chances for promotion or for on-the-job training. If it is the latter, then some part of the lesser market work (and more housework) by women may reflect market discrimination.

(3) *Black women* tend to have a higher U_r and A_r than white women. Again, the research challenge is to determine the extent to which this is attributable to differential discrimination on the demand side compared with the difference in supply-side characteristics between white and black women. Keep in mind that certain supply-side factors, such as the century-long commitment of black women to market work, their lower probability of marriage, and their higher probability of marital dissolution are all plausible reflections of the labor market discrimination faced by black men.²⁵ Thus, the low earnings of black men are in part a cause of the high work rates of black wives and, perhaps of the lower proportion of black adults who are married.

(4) *Adding more control variables* usually raises A_r , and there is a noteworthy pattern to this. Various "pre-labor-market" controls, such as education, age, family background, and residential location, are all very similar for men and women, unlike the case of white men compared to black men. Standardizing the women's predicted earnings with men's mean values for these control variables can hardly close the gap by much. Thus, Blinder shows no difference between U_r and A_r (both equal 0.54) when he holds constant age, health, residence, and family background. Nor would education have made much difference, because the means for men and women in his sample are about the same. These pre-labor-market variables often differ substantially between blacks and whites, however, so controlling for them does raise A_r relative to U_r , as we shall see in Table 13.7.

The variables that reflect work experience, such as the number of years spent in the labor force and the worker's tenure with a firm have, on the other hand, substantially different mean values for men and women. When Blinder added tenure, union status, and a one-digit occupational classification, the A_r rose to 0.70 (from a $U_r = 0.54$). A strong point of Blinder's study is his distinction between variables that are reasonably viewed as being exogenous to the process of labor market discrimination from the variables that are likely to reflect labor market discrimination.

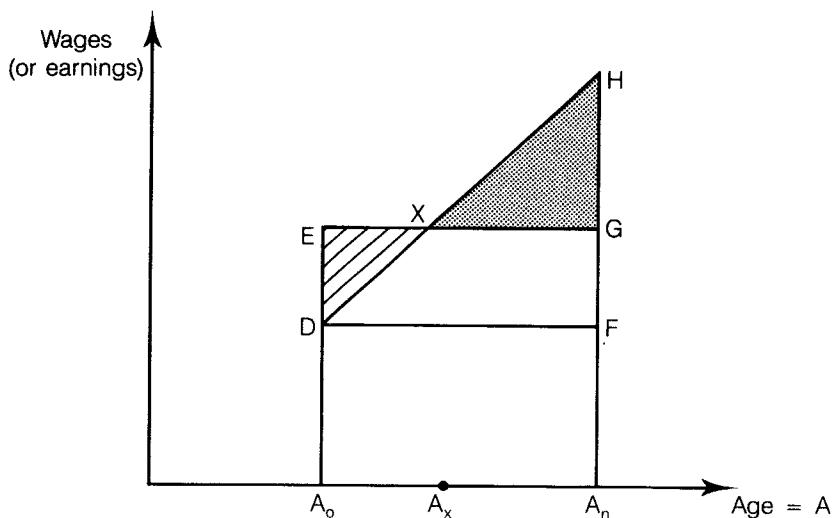
²⁵ The percentage of black women who had ever been married tends to be slightly lower than this percentage among white women, holding age constant. The percentage of black women who were divorced or separated at the time of the surveys is two to three times as large as this percentage among white women, holding age constant. These statistics refer to the years 1970 and 1982. See U.S. Bureau of the Census (1983c, pp. 33, 44-45).

(5) *Using a wage rate as the dependent variable and controlling for years of experience* usually raises A_r , as is illustrated by Corcoran and Duncan and by Mincer and Polacheck (M&P). Both studies measure a relatively narrow definition of discrimination in which the years of experience of the workers are carefully controlled. In these studies the A_r 's rise to 0.80 and 0.85 for married persons and to 0.87 for single persons. These ratios are almost as close to unity as those for which occupation is controlled (Sanborn, Malkiel and Malkiel, Astin and Bayer, and Johnson and Stafford). However, if tenure is a reflection of discrimination – “last hired, first fired” – and if years of experience are less for women because of the lower wage offered to them, then tenure and experience are in the same category as occupation; that is, invalid control variables because they reflect discrimination.

M&P deal with the endogeneity of experience in one of their models by substituting the predicted value of experience in place of actual experience in the wage equation for women. [This technique is also used by Zabalza and Arrufat (1983), who estimate the wage difference between women and men in Great Britain.] The validity of this technique, however, depends crucially on two assumptions. (1) There is at least one variable in the equation predicting experience that is excluded from the equation predicting wages. (2) The excluded variable, which serves to identify the “experience effect” in the wage equation, does not reflect labor market discrimination. The key predictor variable that is excluded from the wage equation and included in the experience equation is the woman's number of children. Are the above two assumptions satisfied? The question is debatable, but I believe the presence and number of children shifts the issue of discrimination onto another dimension of what are simultaneously determined behavioral outcomes: time in market work, time in housework, numbers of children, occupational choices and career plans, and so on.

Polacheck (1979) does treat experience and occupation similarly, because he views both as simultaneously chosen by women in view of their greater commitment to housework and their lesser commitment to market work compared to men. Polacheck argues that women will choose occupations that facilitate their intended short and intermittent stays in the labor market; specifically, occupations with relatively flat age–earnings profiles that do not offer the large rewards to experience as do occupations that provide relatively steeply rising age–earnings profiles and which tend to be male-dominated. The theory of Polacheck and M&P of the time allocations to work over the life cycle offers an explanation for why market experience is less for women and also for why women's wage returns to experience are less – that is, why their age–earnings profile is flatter.

Figure 13.2 clarifies these ideas. Consider the three age–earnings paths, DF , EG , and DH , drawn linearly to simplify the exposition. Equally productive workers, who start at age A_0 and who retire no later than A_n , may be assumed to be indifferent between occupations with the age–earnings profile DH or EG because, let us assume, the present values of the two streams of earnings are the



A_0 = Age at which worker enters labor force

A_n = Retirement age

EG, DH = Two age-earnings profiles chosen by equally productive workers

DF = Age-earnings profile chosen by a worker who has chosen to invest less in earnings capacity (or human capital)

Figure 13.2. Three hypothetical age-earnings profiles. A_0 = age at which worker enters labor force; A_n = retirement age; EG, DH = two age-earnings profiles chosen by equally productive workers; DF = age-earnings profile chosen by a worker who has chosen to invest less in earnings capacity (or human capital).

same. The cross-hatched area, DEX , is drawn to be smaller than XGH to allow for the discounting of future receipts.

If the equally productive workers are men and women, the M&P theory is that women will choose EG instead of DH because they are less likely to want to stay in their job continually (from ages A_0 to A_n), and they will prefer the higher wages in EG up to age A_x instead of choosing the DH path. This part of the M&P theory would not, by itself, be supported empirically: we do not observe women earning higher wages than men between ages A_0 to A_x . A second part of the theory offers a supply and demand explanation for why women's profiles will

be like *DF* instead of *EG*. On the demand side the claim is that employers generally prefer workers who are willing to work continually and who are willing to accept the *DH* profiles (and the on-the-job investment that *DH* implies). This decreases the demand for the *EG* workers and lowers their earnings profile. On the supply side, the choice by women to work intermittently implies that they will not invest as much in marketable human capital, because they will have fewer years to receive returns on their investment. This lowers the earnings path still further, say to *DF*.

Thus, the supply-side argument for lower earnings paths for women is that they invest less. However, when the investment is measured by years of schooling, we do not observe important differences between men and women. The argument sometimes shifts to an emphasis on less observable variables, such as the intensity of investment in schooling or in on-the-job training. See the comments by June O'Neill, below.

In principle, these sorts of assumptions about gender behavior can rationalize an A_r of 0.85 and explain away the remaining gap of 0.15 by references to "measurement error" and other sources of omitted productivity variables that, if corrected, would show men to be more productive. In summary, the argument is that the lesser investment, lesser experience, greater time in housework, and lower occupational attainments of women (a) are voluntary choices made by women and (b) are choices that causally precede the gender difference in the demand structure. But because our economic theories and statistical techniques cannot tell us what is or is not voluntary, I doubt that the computations of A_r 's will measure labor market discrimination in any fundamental sense.

Perhaps the most emphatic argument in support of the "voluntary view" is presented in a series of papers by O'Neill (1983a, 1983b, 1984). Her discussion of the lower occupational status of women is replete with references to choices (from 1983b):

The investment component of schooling ... varies by subject matter. Women have traditionally chosen majors such as education, arts, and humanities, which have lower pecuniary returns than subjects such as business or science (p. 19).

Since many women continue to be responsible for a disproportionate share of household maintenance and child care even after they enter employment, they are likely to evaluate certain job characteristics differently than men...[so] predominantly female occupations were much more likely to offer part-time work and less likely to require very long work weeks (p. 19).

... there is a strong element of personal choice in the occupations held by women... the dominant variables explaining whether a woman is in a typically female occupation were those describing plans and expectations held five years

earlier. Women who said they planned to be a homemaker at age 35, who had children, were married, and who did not attend college, were more likely to be in stereotypically female occupations (p. 22).

Women who five years earlier said they planned to work at age 35 and desired to be working in a male-typed occupation, who attended college, and majored in a scientific subject, were in fact highly likely to be in male-dominated occupations. These findings would appear to contradict the presumption that barriers to entry are the primary reason why women are poorly represented in many occupations (p. 22).

Notice that none of these arguments, which emphasize choice rather than discrimination, is persuasive if discrimination is believed to be causal to the choice of majors, to the time devoted to housework, to the employers' offers of part- versus full-time jobs, and to the "plans and expectations" women had at ages 19–29 regarding their career at age 35. In other words, if labor market discrimination does restrict the quality of jobs and wages available to women, it is reasonable to believe that this affects their plans and expectations regarding school majors, fertility, and their time allocation to home and market sectors. The last quote by O'Neill seems to say that, for example, (a) if a woman who received a degree in electrical engineering is working as an electrical engineer, then (b) the poor representation of women in electrical engineering is not to be considered evidence for discrimination against women ("barriers to entry") in that occupation. It seems to me that (b) does not follow from (a).

(6) *Restricting the sample to unmarried women and men* usually shows higher values of U_r and A_r , as is illustrated by the ratios 0.86–0.87 reported by M&P. Studies that compare single women to either single men or all men might be viewed as providing a purer measure of gender discrimination by avoiding the troublesome issue of the dual career that is associated with married women. Unfortunately, the issue remains. The never-married single women tend to be young, under 25 or so, and the U_r for young people is relatively high (see the Kohen and Roderick entry in Table 13.6). However, a ratio that is less than 1 may reflect the employers' expectations that the women are likely to marry and to be less committed to their jobs than men of the same age. If the sample were restricted to never-married women in their 40s or older, for whom a strong commitment to market work may be presumed, the sample would be relatively small and probably selective of women who were either unusually dedicated to a career or unusually adverse to marriage. Arguments could be made that these women would be likely to earn more, or less, than men who are comparable in the conventional characteristics used in earnings functions. Indeed, single men tend to earn less than married men, holding constant conventional variables. A full understanding of these selective traits determining marital status involves more than just economics.

(7) *Restricting the comparison to a narrowly selected group of jobs* tends to produce higher U_r and A_r ratios, as is illustrated by Malkiel and Malkiel (1973). This study is the only one in Table 13.6 for a single company, and I will have more to say about this type of sampling restriction later. Also, it is not only a sample of a relatively narrowly defined occupational group – all college-educated professionals who work for a particular research firm – but it offers a control over “job level”, which further narrowly defines the tasks, duties, and responsibilities of the employees. By controlling for job level, the adjusted ratio rises from 0.77 to 0.86. This is evidence for the claim made earlier that with a sufficiently narrowly defined job almost all ratios would be unity. Indeed, if not, companies would risk violating the law. Finally, the study offers a rather striking example of the importance of the method of standardization. The conventional A_r is equal to \hat{W}_f/\bar{W}_m , and an alternative is $A'_r = \bar{W}_f/\hat{W}_m$. The alternative adjusted ratios in the Malkiel and Malkiel study are 0.85 (instead of 0.77) and 0.99 (instead of 0.86). The 0.99 ratio was the one emphasized by Malkiel and Malkiel and used by O’Neill (1984, pp. 79–82).

4.2.2.2. Black–white earnings gap for men. The empirical measurements of wage discrimination between blacks and whites in the United States involve the same procedures as those just described for men and women, but there are differences in results and interpretations. I focus on men in the racial comparison in order to separate this from the gender factor. A difference in the normative interpretation is that the wage ratios for women-to-men, U_r and A_r , that are less than one may be rationalized by claiming that women choose to specialize in home production. No such alternative employment is credible for black men. Furthermore, even if women suffered lower market wages because of discrimination, they might recover all or part of these losses by marrying men, who are the favored group. Consideration of total household income for comparisons between men and women will be discussed in the last section.

The unadjusted wage ratios for black and white men shown in Table 13.7 are similar to those for women-to-men in Table 13.6, if we restrict the comparisons to large populations and exclude the comparisons for young people and for selected occupations. The average U_r is 0.58 in the black-to-white male ratios and 0.55 for women-to-men. The increase in the adjusted ratios, $A_r - U_r$, is, however, generally larger for the race ratios in Table 13.7 than those for the gender ratios in Table 13.6, if we exclude the comparisons with asterisks where occupation is held constant. The average increase in A_r relative to U_r is about 0.16 (0.74–0.58) for men, black-to-white, and about 0.09 (0.65–0.56) for women-to-men. Despite the crudity of these comparisons – they are not confined to comparisons with similar variables held constant, for example – we may conclude that holding constant the usual available productivity variables has a larger effect in reducing the wage gap between black and white men. The reason, mentioned earlier, is that the usual

Table 13.7

Summary of studies of ratios of black men's earnings to white men's earnings, unadjusted and adjusted for various characteristics of workers and jobs.

Author and year of publication ^a	Data source and population studied ^b	Measure of earnings ^c	Statistical method and explanatory variables ^d	Blacks' earnings as a ratio of whites'	
				Observed ^e	Adjusted ^f
Masters (1975)	Census (1/1000 sample), blacks, non-Hispanic whites, age 17-64, worked in 1959, civilian, nonstudent	y, 1959	R, S: 1, 2, 12, 13, 25	0.50	0.59
			above + 10 ^g	0.50	0.80
Blinder (1973)	PSID, white and black working household heads	w, 1969	R, S: 1, 2, 12 13, 25	0.64	0.61
			above + 10 ^g R, S: 2, 9, 12-14, 32, 34	0.64 0.49	0.74 0.64
Masters (1975)	SEO, blacks, non-Hispanic whites, age 17-64, worked in 1966, civilian, nonstudent	y, 1966	R, S: 1, 2, 12, 13, 25	0.55	0.66
			above + 10 ^g	0.55	0.72
Same as above but for non-South			R, S: 1, 2, 12, 13, 25	0.68	0.75
Duncan (1968)	OOC, native, nonfarm, 25-64	y, 1961	R, S: 1, 4, 8	0.46	0.68
Corcoran and Duncan (1979)	PSID, household heads, worked 500 hours or more, age 18-64	w, 1975	above + 4 ^g	0.46	0.75
			R, S: 1, (2), 5, 6, 9, 11-13, 16, 17	0.77	0.89
Flanagan (1974)	NLS, men 15-25 or 46-60	w, 1967	R, S: 1, (2), 5, 6, 8, 9, 11, 13, 26, 35	0.58	0.84
			Age 46-60	0.58	0.84
			Age 15-25	0.72	0.94

Sources: Full citations of the original studies are in the bibliography. In some cases the adjusted ratios that are shown in this table do not appear in the original studies or appear in a different form.

Note: All footnotes are identical to the footnotes in Table 13.6.

control variables generally represent either exogenous characteristics, like age, or pre-labor-market characteristics, like years of schooling, and these types of variables tend to be more similar for men and women than they are for whites and blacks.

Blinder's (1973) comparisons are again instructive. His control for exogenous variables made no difference in the women/men ratios; both U_r and A_r equal 0.54, but the same control variables raised the black/white ratios from 0.49 to 0.64.

The 1979 study by Corcoran and Duncan (C&D) is also interesting, and some comments about it serve to raise several general points. The unadjusted ratio,

$U_r = 0.77$, is the highest in Table 13.7. C&D use a wage rate rather than earnings as the dependent variable, and they exclude workers who worked less than 500 hours during the survey year. Both restrictions raise the black/white ratio, because black men are likely to suffer more unemployment, including unemployment for 10 months or longer. Recall that restricting women/men comparisons to full-time workers (or controlling for hours worked) was primarily justified on grounds that the frequency of part-time jobs among women was often voluntary. This is seldom true among prime-age black men.

C&D also restrict their sample to men who are household heads, and there is likely to be some selectivity bias here that raises the wage ratio above what it would be for the full population of black and white male workers. A smaller proportion of black men aged 18–64 are household heads compared to white men of these ages; the wages of those who are not household heads in both races are lower than the wages of heads of households; the wage ratio of blacks-to-whites among men who are not household heads is slightly less than 0.77; and household headship may reflect labor market discrimination.²⁶ Therefore, the black-to-white wage ratio for all men would be less than 0.77.

The adjusted ratio in the C&D study is 0.89, also high relative to other A_r 's. The control variables include years of experience which, given a control for years of schooling, is representing age and therefore almost purely exogenous. Let us assume, for the sake of argument, that years of schooling, city size, region of residence, and health, are exogenous. The remaining control variables, formal training and tenure with one's current employer are, however, likely to reflect labor market discrimination.

The ratios by Blinder and by C&D are virtually the lower and upper bounds in Table 13.7. The other studies suggest several additional methodological issues, but I will be brief. Masters' studies (1975) show the importance of the South/non-South differential. More recent data in the C&D study show that this differential is still important, although smaller. Masters clearly brings out the effect of controlling for time worked, because each of the A_r 's for the "above +10" comparisons (see the fourth and sixth columns) allows only the additional control for weeks worked, and these A_r 's are much larger. Finally, although I do not show them, there are some striking differences between the A_r and A'_r ratios in Masters' study.

²⁶Among men aged 18–64, which is the population frame for Corcoran and Duncan, 79 percent of white men and 58 percent of black men were household heads at the time of the Census Bureau's survey in 1981 [U.S. Bureau of the Census (1983a, P-20, No. 372, Table 2)]. The incomes of year-round full-time workers who are male heads of households is about 25 percent higher than for similar workers who are not heads of households. The ratio of black-to-white income for men who were not household heads in 1981 is around 0.70 for all workers and 0.76 for full-time workers [U.S. Bureau of the Census (1983b, P-60, No. 137, Tables 44 and 55)]. Income figures are used instead of earnings because earnings are not reported for persons classified by their relationship to the household head.

The study by the sociologist Otis Dudley Duncan (1968) may have been the first to use separate regressions and to construct the "decompositions" of wage (or income) differences. Economists usually cite later studies by economists for these procedures.

Flanagan's studies show the frequently observed result that black/white wage ratios are relatively high for young workers, which was also true for the gender ratios. However, the smaller ratio among older men surely does not reflect a voluntary choice by black men to work less in market employment, as might be claimed for women. On the other hand, so-called "vintage effects" may be revealed in the different wage ratios for younger and older black and white groups, whereby the current period's larger ratio for young workers may reflect a true long-run improvement in the relative earnings capacities of black men—perhaps reflecting, in turn, recent improvements in the quality and quantity of education. Welch (1973) and Smith and Welch (1977, 1978) have stressed this source of a vintage effect. Others have emphasized the civil rights movement in the last 20 years, a reduction in discrimination in society, and the increase in antidiscriminatory legislation, all of which may be having a larger positive effect on young blacks than on older blacks [see Freeman (1981)].

Clearly, current wages of young blacks and whites cannot conclusively reveal a lifetime comparison. Cohort analyses of previous generations show that only part of the improvement in wage ratios among previous generations of young people is sustained. See Freeman (1973a), Chiswick (1974, pp. 116–118), Smith and Welch (1978), and Hoffman (1979).

4.2.2.3. The earnings gap for other ethnic groups. The nationality group in the United States that has received the most attention in discrimination studies in recent years is Hispanics, which consists predominantly of persons with Mexican, Puerto Rican, and Cuban ancestry, in that order. There is not the space to review the empirical estimates of Model (I) for these groups and to display a corresponding table. Nevertheless, some new and interesting methodological issues may be mentioned briefly in connection with the general finding that relatively small unadjusted wage ratios (U_r), Hispanic-to-white, coexist with relatively large adjusted ratios (A_r). ("White" refers to non-Hispanic white.) For example, Reimers (1983) finds that U_r for Mexican Americans (hereafter Mexicans) relative to whites is about 0.70, whereas the adjusted ratio after fitting separate Model (I)-type regressions is about 0.94. See also Abowd and Killingsworth (1982) and Grenier (1984) for similar results. Adjusted ratios of 0.9 or higher imply a minor role for labor market discrimination.

There are four main sources for the increases from U_r to A_r in these studies: (1) Age differences: Hispanics tend to be younger than whites, so part of the wage difference is explained by this exogenous variable. (2) Education differences: Hispanics, particularly Mexicans and Puerto Ricans, have substantially lower

average educational levels. (3) Immigration status, as measured by years in the United States. (4) Language, as measured, for example, by a categorical variable defined as the ability to speak and understand English. All these controls seem reasonable, but let us critically examine the last three and raise again the point that definitive empirical measures of wage discrimination require knowledge about the underlying processes and institutions.

Education. Assume that education contributes to one's earnings capacity in skilled jobs but not in unskilled jobs. If this were true, then the fact that, say, Mexicans have a low average years-of-schooling-completed (ED) should not place them at an earnings disadvantage relative to whites in the unskilled jobs, even though whites in these jobs may have a larger mean ED , perhaps because of school attendance laws in the U.S. Nevertheless, the usual regression procedure, with ED entered linearly in the earnings regression, will tend to assign a lower predicted wage to Mexicans relative to whites in these unskilled jobs. Actual discrimination in these jobs may be masked. The following simple hypothetical example brings out this point. Assume that only ED , among the observable variables, has a systematic effect on wages, and that the distribution of wages (W) and ED is as shown in Panel A of Table 13.8. There are two examples each of distributions for whites and Hispanics, which permit four possible calculations of U_r and A_r . To obtain A_r with the data of example 1, we obtain the average predicted wages for Hispanics and whites:

$$\hat{W}_w = -0.46 + 0.357ED, \text{ so } \bar{W} = -0.46 + 0.357(\bar{ED} = 11.93) = 3.80 \quad (12)$$

and

$$\hat{W}_h = -0.60 + 0.375ED, \text{ so } \bar{W} = -0.60 + 0.375(\bar{ED} = 8) = 2.40. \quad (13)$$

The corresponding regression equations using example 2 data are

$$\hat{W}_w = 0.025 + 0.322ED, \text{ so } \bar{W} = 0.025 + 0.322(\bar{ED} = 11.73) = 3.80 \quad (14)$$

and

$$\hat{W}_h = -0.300 + 0.313ED, \text{ so } \bar{W} = -0.300 + 0.313(\bar{ED} = 8) = 2.20. \quad (15)$$

The A_r and A'_r formulas, using (12)–(15), give us the A_r entries in Panel B of Table 13.8.

The data have been constructed to reveal wage discrimination against Hispanics that is *assumed* among unskilled workers, who are all classified with low levels of ED – an average of 4 for Hispanics and an average of either 4 or 5 for whites. In example 1 of Hispanic data there is no discrimination among the higher ED

Table 13.8
Hypothetical distribution of years of schooling (ED) and wages (W) among whites and Hispanics
and resulting comparisons of unadjusted and adjusted wage ratios

Years of schooling (ED)	Panel A. Distribution							
	Whites				Hispanics			
	Example 1		Example 2		Example 1		Example 2	
	Number	Wage	Number	Wage	Number	Wage	Number	Wage
4	1	\$2	3	\$2	3	\$1	3	\$1
5	1	2						
6	1	2						
12	7	3	7	3	1	3	1	3
16	5	6	5	6	1	6	1	5
Number	15		15		5		5	
Means	$\bar{ED} = 11.93$		$\bar{ED} = 11.73$		$\bar{ED} = 8$		$\bar{ED} = 8$	
	$\bar{W} = \$3.80$		$\bar{W} = \$3.80$		$\bar{W} = \$2.40$		$\bar{W} = \$2.20$	

Panel B: Comparisons of unadjusted (U_r) and adjusted (F_h, A_r, A'_r) wage ratios*								
Hispanic distribution								
White distribution	Example 1				Example 2			
	U_r	F_h	A_r	A'_r	U_r	F_h	A_r	A'_r
Example 1	0.63		1.02	1.00	0.58		1.02	0.94
Example 2	0.63	0.70	0.99	0.93	0.58	0.67	0.88	0.85

Source: Hypothetical numbers.

Definitions and explanations of ratios in Panel B:

$$U_r = \bar{W}_h / \bar{W}_w.$$

$$A_r = \bar{W}_h / \bar{W}_w = \sum b_h \bar{X}_w / \bar{W}_w.$$

$$A'_r = \bar{W}_h / \bar{W}_w = \bar{W}_h / \sum b_w \bar{X}_h.$$

F_h for example 2 of whites with example 1 of Hispanics:

$$F_h = (3/5)(1/2) + (1/5)(1) + (1/5)(1) = 0.70.$$

F_h for example 2 of whites with example 2 of Hispanics:

$$F_h = (3/5)(1/2) + (1/5)(1) + (1/5)(5/6) = 0.67.$$

categories, and in example 2 there is discrimination for the college ($ED = 16$) category—a wage ratio of $5/6 = 0.83$.

The unadjusted wage ratios with the data reveal typical values: 0.63 and 0.58. An exact measure of wage discrimination, given the assumptions used in constructing the data, is available with example 2 of the white data and both sets of the Hispanic data. This measure, expressed as a ratio, will be labeled F_h to indicate that the ratio is based on a “free functional form” (using categorical variables for each level of ED) and that it uses Hispanic numbers of workers as weights. It is calculated as follows: With example 2 of the white data, we see that 3 of the 5 Hispanics have the same ED as whites and earn only half as much. With example 1 of the Hispanic data, the other two Hispanic workers with higher ED earn the same as whites with the same ED values. F_h is the Hispanic/white

wage ratio, adjusted for ED values in each of the three ED categories.

$$F_h = (3/5)(1/2) + (1/5)(3/3) + (1/5)(6/6) = 0.70.$$

No functional form has been imposed on the ED/W relation, and the wage comparisons at each ED level are weighted by the relative frequencies in the Hispanic distribution.²⁷

The main conclusions from Table 13.8 are that the A_r and A'_r ratios show either minimal or no discrimination, despite its “constructed” presence, and that the sources of the discrepancy are the incorrect linear relation between ED and W and the disparity in numbers of observations in the categories – few Hispanics in the higher ED categories and relatively few whites in the low ED category. The example is oversimplified, of course, but the problem it reveals with the typical econometric estimation for Hispanics, particularly Mexicans, is, I believe, worth considering. The general issue is that a trait, in this case ED , may vary in its validity as an indicator of productivity over certain ranges of the variable and in certain work situations.

Immigration. Years since arrival in the United States may be an indicator of productivity, because the variable may represent English-language skills, labor market information (including investments in job mobility), and various aspects of cultural assimilation that may enhance workers’ productivity in their contacts with supervisors, co-workers, customers, and so on. If the lack of cultural assimilation has nothing to do with one’s productivity and everything to do with discrimination, the variable loses its appeal as a control variable on this account, although it retains its appeal on the other two accounts.

Language skills. As a control variable, language skills may be partially contaminated by simultaneity if the skills depend on the type of job available to the worker, on the relationship to one’s co-workers, and on other outcomes of the labor market. In one study the variable was highly correlated with “place of birth” and “number of years spent in the U.S.” and the latter variables were omitted from the regression. The author [Grenier (1984, p. 42)] commented: “One consequence of these omissions, however, is that estimated coefficients of the language variables may include some effects of these other variables.”

In addition to the study and measurement of discrimination regarding Hispanics, there have been a few econometric and historical-institutional studies

²⁷ Using white relative frequencies, we have $F_w = (3/15)(1/2) + (7/15)(3/3) + (5/15)(6/6) = 0.80$, but even this moderate measure of discrimination is not appropriate to describe the situation of Hispanics in this hypothetical example, where 60 percent (3 of 5) are experiencing a large measure of discrimination. A more appropriate use of the F -ratio for whites is to define

$$F'_w = (3/15)(2/1) + (7/15)(3/3) + (5/15)(6/6) = 1.20,$$

which shows that the large (two-to-one) wage advantage of the unskilled whites raises their overall wages relative to Hispanics by 20 percent, holding education constant. Finally, if we assume, with example 1 data for whites, that the ED levels of 4, 5, and 6 provide no productivity differences among these workers, then F_h and the other F -ratios would be defined exactly as they are for the example 2 data for whites.

of several European and Asian nationality groups and various religious groups. One important result of these studies is that for many ethnic groups, the ratio of their earnings to those for a more broadly defined white group (sometimes as narrow as those with an English ancestry) is larger than one. This was found, for example, for Catholic Irish-Americans [Greeley (1976, p. 52), and (1981, pp. 110–120)], Japanese-Americans [Petersen (1978), Sowell (1981)], and Jews [Chiswick (1983), Sowell (1981)]. The current advantaged status of these groups has been explained by particular historical and institutional developments, rather than as revealing “reverse discrimination”, and these explanations persuade me of the value of this method of analysis.

In a study of the relation between the larger religious groups and earnings, Tomes (1984) found small and statistically insignificant effects among Catholic, Protestant, and “None/Other” categories.²⁸ He also found no statistically significant difference among various Protestant denominations. Tomes provides a useful distinction between estimates with purely exogenous variables held constant – family background, age (and age squared), and location of residence – and estimates in which potential “outcome” variables, like education and self-employment, were additional control variables. Another interesting feature of this study was the distinction between one’s current religious affiliation and that of one’s upbringing. Current affiliation is, to some extent, endogenous, and there were some interesting, although not startling, differences in the estimates when the two definitions of religious affiliation were used.

4.2.3. Empirical studies of wage discrimination in individual firms and “reverse regression”

A brief examination of the econometric analysis of discrimination in individual firms is useful for three reasons. First, discrimination by firms, identified singly, has come under close scrutiny by various groups in society, mainly as a result of antidiscrimination laws and executive orders. Regression analyses of the Model (I) type are frequently offered as evidence in court cases and other litigation proceedings stemming from these laws and regulations.²⁹ Discrimination in firms may consist of the differential treatment of majority and minority groups in hiring, placement, wages, promotion, layoffs, and in other ways. I will refer only to wage discrimination in this discussion.

A second reason for interest in these studies is that they have the advantages, relative to marketwide studies, of explicit and well-defined objectives and

²⁸ Tomes found relatively large positive effects for Jews, but this group was numerically small in his sample and the differences were sometimes not statistically significant.

²⁹ See Baldus and Cole (1980), Finkelstein (1980), and Fisher (1980) for a discussion of statistical analyses in court cases of discrimination and for extensive citations. An example of an econometric study of wage discrimination in a single firm, listed in Table 13.6, is that by Malkiel and Malkiel (1973), although this study was not used in litigation proceedings.

straightforward procedures. The analysts' objectives are usually to support their clients, and the objective of the court is to use these studies to determine whether an employer is guilty or innocent. The procedures involve a Model (I)-type regression in which the employer's criteria for wage payments may be specified in detail, and those criteria that are correlated with race or gender can be explicitly included in the estimation model. Moreover, many characteristics associated with the workers' productivity may be clearly exogenous to the employer, even though they are not exogenous to the market as a whole. Regression analyses with nationwide samples usually suffer from ambiguities and vagueness about both their objectives and model specifications.

A third reason for our attention is the innovative use of "reverse regression", which raises some interesting methodological points even if, as I will argue, it does not offer a preferred model for estimating wage discrimination. Indeed, this section will examine only this aspect of the studies of individual firms. The actual analyses are often buried in trial proceedings, and there is no space to present them here. More important, I believe that studies of discrimination by single firms do not provide useful measures of marketwide discrimination, which is my main interest. The reason is that the samples are based on company records, and the selection rules for inclusion in the sample are seldom known. The companies studied in court cases are not a random sample of all companies, and their recruitment policies do not pretend to yield random samples.

Company records generally apply to a single industry and a few occupations, and the question of how market discrimination affects the distribution of minority workers in the industry and occupation is not examined. More generally, no valid conclusions about discrimination can be reached without attention to the company's recruitment or selection procedures. For example, perhaps the company has a reputation for discrimination against women that restricts the pool of female applicants. Maybe only a small number of newcomers to the community constitute this pool. The statistical analyst usually deals with the employees on board or, at best, with persons who have applied to the company. A famous study by Conway and Roberts (1983) of alleged salary discrimination against women in a large bank, for example, involved a sample of 274 employees, of whom only 37 are women. Under these circumstances, generalizing about discrimination in the market as a whole on the basis of studies of one or several companies is not valid.

Studies by Roberts (1980) and by Conway and Roberts have given prominence to the statistical model known as "reverse regression" (hereafter RR).³⁰ The term refers to "reversing" the Model (I) regression in which the wage, W , is regressed

³⁰A large number of articles on reverse regression in discrimination analyses appeared around the time and soon after Roberts's article (1980), and a symposium on the issue appears in *Journal of Business and Economic Statistics*, vol. 1, January 1983. My understanding of the issues owes much to Arthur S. Goldberger, and my discussion is based on Goldberger (1984), but I am solely responsible for any errors in this section.

on a vector of productivity variables, X , and a minority/majority-status variable, Z . In RR , X is in effect regressed on W and Z . Assume Z refers to gender. One motivation for RR may lie in the question it addresses: "Holding wages constant, are men less qualified than women?" This reverses the customary question in direct regression: "Holding qualifications (the X 's) constant, are men paid more than women?"

Another motivation for RR is the classic problem, in regression analysis, of errors in the independent variables. Given the regression, $\hat{W} = a + bX$, in which X contains random measurement error but wherein the regression is otherwise correctly specified, RR —here, $\hat{X} = c + dW$ —permits the coefficients b and $1/d$ to give boundaries on the true linear relation between W and the correctly measured X . When X is a collection of variables, the dependent variable in RR may either be (a) each individual X -component regressed against W and Z in a system of equations, or (b) the fitted part of the direct regression for the case in which $Z = 0$, specifically, $\hat{W} = X'B$. Thus, $X'B$ is regressed on W and Z in RR . I will sometimes simply refer to X as the "dependent variable" in RR .

The initial appeal of RR is the reasonable proposition that the econometrician's usual set of productivity variables, X , is not a perfect measure of a worker's productivity. Assuming that W is a function of "true productivity", X^* , the following model may be specified:

$$W = X^* + AZ + v, \quad \text{where } Z = 1 \text{ for men and } Z = 0 \text{ for women;} \quad (16)$$

$$X^* = GZ + u, \quad \text{where } G > 0; \quad (17)$$

$$X = X^* + e. \quad (18)$$

By eq. (16) the wage is a function of true productivity, X^* , which is unobserved by the econometrician. Holding X^* constant, $A = 0$ implies no discrimination against women, and $A > 0$ implies discrimination against women. Equation (17) says that men are more productive than women; this assumption will be maintained throughout this discussion. Equation (18) says that X is a fallible measure of X^* . The usual assumptions about the error terms are that v , u , and e are independent of each other; that v and e are independent of X^* and Z ; and that u is independent of Z .

Assume that the econometrician estimates the direct regression:

$$W = BX + CZ + v'. \quad (19)$$

It follows from (16)–(18) that C is an upwardly biased estimate of the true

relation, A . Assuming $A = 0$, C will be greater than zero, implying discrimination against women when none exists. [The algebra and further discussion of these qualitative results are found in Goldberger (1984).] In eq. (19), X^* is an omitted variable, as revealed by the assumed correct wage eq. (16). By (17), X^* is systematically related to gender. Thus, positive values of Z in (19) partly represent "more" X^* in addition to representing "maleness" (since $Z=1$ for men). Therefore, the coefficient C will be positive owing to the relation of X^* to Z even if the true male effect, A , is zero. Clearly, C is more upwardly biased as G is larger in (17) and as the variance of e is larger (implying that X is a less accurate measure of X^*) in (18).

At first glance, eqs. (16)–(18) may seem so plausible that the systematic upward bias of C in (19) seems incontrovertible. The first glance is misleading. There is no basis for assuming that the employer pays according to "true productivity", X^* . On the contrary, we should expect that the systematic basis on which employers pay their workers is a basis of observable variables. A random error term should also be included to allow for miscellaneous factors that may be assumed unrelated to gender. This latter point reintroduces the earlier argument that in a discrimination case, any systematic productivity trait that is correlated with minority-group status should be included in the employer's list of the X -variables. These assumptions effectively restore to legitimacy either eq. (19), for the case of a single X -variable, or eq. (20), as written below, with multiple X -variables:

$$W = X' \mathbf{B} + CZ + v'. \quad (20)$$

We may further assume that there is generally a positive relation between X and Z in the sense that the men's mean of the X -vector, weighted by their "prices" – the \mathbf{B} -vector – is larger than that for women, that is

$$E(X'_m) \mathbf{B} > E(X'_w) \mathbf{B}. \quad (21)$$

This inequality is the analogue of (17), which expressed the assumption that men are more productive than women. Given the assumption that the X 's are positively correlated with X^* , (21) implies $G > 0$ in (17), and conversely. However, (17) and (21) do not imply that the expected value of X^* , holding constant the observable X 's, is greater for men than for women, and without this assumption there is no basis for assuming that C is upwardly biased in (20).

Despite the assumed correctness of (17) and (21), the direct regression of (20) gives an unbiased estimate of the gender (male) coefficient. Equation (20) has the virtue of focusing attention on the explicit measure of alleged discrimination, W , and of leading all the interested parties – the econometrician, the defendant employer, the plaintiff, and the adjudicator – to address the same questions raised

earlier in this chapter about Model (I):

(a) Are the variables in X proper control variables; that is, are they exogenous with respect to the employer's behavior?

(b) Are there omitted productivity variables that are systematically related to gender?

It should be clear that eq. (20), which expresses an observable relation and is to be analyzed in conjunction with an inquiry into questions (a) and (b), allows for the model expressed as eq. (16) to be a special case. As mentioned earlier, X^* could be assumed to be an omitted variable and, given (17)–(18), the upward bias in C follows. But there is no reason to believe that any such concept as "true productivity" will be operational. Furthermore, if X^* is redefined to be the employer's "assessed productivity" [see Roberts (1980), Conway and Roberts (1983)], then we return to the reasonable expectation (or requirement) that the employers simply point out which criteria in their assessment have a net correlation with gender—that is, a correlation that persists after controlling for the observable X 's. Again, a straightforward analysis of "omitted variable" bias in eq. (20) could provide a qualitative answer about the direction of bias in C . The direct regression associated with (20) leads us to focus attention on specific sources of any bias. This has more appeal to me than a model in which the wage is presumed to be determined by "true" (or "assessed") productivity combined with the presumption that male superiority regarding that productivity is maintained after holding constant observable measures of productivity. These presumptions prejudge the very issue that is to be investigated, and these presumptions are precisely what eqs. (16)–(18) embody.

The model in eqs. (16)–(18) has two multivariate analogues that have been analyzed by Goldberger (1984). In one, each X -variable in the vector of observed indicators, X , is assumed to be a fallible measure of the corresponding element in a vector of unobserved true productivity determinants, X^* . In this case RR is not necessarily superior to direct regression even though one may assume, analogous to (17), that

$$X^* = HZ + u, \quad (17a)$$

with all elements of $H > 0$.

In the other multivariate model the X -vector is assumed to provide multiple indicators of a single X^* , so (18) becomes

$$X = DX^* + e. \quad (18a)$$

Let a representative indicator be X_j . The RR for the j -indicator is

$$\hat{X}_j = c_j + d_j W + f_j Z, \quad (18b)$$

and the estimator, A^* , of the gender coefficient in the true wage equation, (16), is $A_j^* = -f_j/d_j$.

This multiple-indicator model is said to be the one “stochastic specification under which reverse regression provides a valid estimator” of the gender coefficient [Goldberger (1984, p. 314)]. Because each indicator of X is assumed to provide a consistent estimator of the true gender coefficient, these implied restrictions on the different estimators can be tested. Few such tests have been tried, and those few are hardly supportive of the model [see Goldberger (1984), Green and Ferber (1984)].

One of the questionable assumptions of the multiple-indicator model as it applies to gender discrimination is that each indicator, X_j in X , is independent of gender, holding constant X^* . This assumption is associated with eq. (18a), where e is assumed to be independent of gender, or with eq. (18) in the classical errors-in-variable model with one independent variable. To illustrate why this assumption and the implied restrictions on the coefficients in (18b) are not likely to hold, consider an X that has the following four elements, which are positively correlated with true (or assessed) productivity, X^* , but which differ in their correlation with gender: X_1 = verbal test score; X_2 = mathematics test score; X_3 = manual dexterity; X_4 = physical strength.

The gender correlations will reflect, let us assume, an advantage of women over men in verbal ability and manual dexterity and an advantage of men over women in mathematical ability and physical strength, which are gender-linked relationships that have been generally found [see Anastasi (1969, p. 421)]. To keep the statistical relations simple, assume that each element of D is the same. Then if a nondiscriminating employer hires men and women at a specified wage, we should observe that women’s low scores on X_2 and X_4 will tend to be compensated by higher scores on X_1 and X_3 , and conversely for men. The dependence of e with gender ($= Z$) in (18a) will induce different signs of the f_j in (18b). In particular, with $Z = 1$ for men, f_1 and f_3 will tend to be negative and f_2 and f_4 will tend to be positive. With d_j positive, the estimators of the true gender effect, A_j^* , will not all have the same sign.

The reasons why the multiple-indicator model is invalid in this example are, I believe, realistic and prevalent; specifically, the presence of some gender-linked productivity traits that favor women, the fact that employee productivity typically involves multiple skills, and the fact that employers recognize that skills (and traits) are typically compensating (or substitutable).³¹

³¹A hypothetical example of substitutable gender-linked traits that brings out these statistical points is available from the author. It should be noted that a model in which wages are determined by gender-linked traits, X , along with an assumption that other traits, holding constant X , are uncorrelated with gender simply illustrates the model in eq. (20), which is the same as the “multiple cause model” that Goldberger (1984) discusses.

5. Policy implications and conclusions

5.1. *Explaining and judging discrimination: The diversity of cases*

This chapter began with the normative issue of equity in outcomes measuring economic well-being among racial, ethnic, and gender groups. Inequities appear to be widespread, and our economic theories of why they persist are only moderately helpful.

At one extreme, the outcomes experienced by earlier immigrants to the United States suggest an optimistic view of both the ethical and the scientific judgments about the workings of the economy generally and of labor markets more particularly.³² Although discrimination against early immigrant groups was not analyzed in this chapter, the references to the achievements of immigrants who were Irish Catholics, Italians (mainly from southern Italy), Japanese, and Jews (mainly from eastern Europe) seem to show a pattern in which groups who were initially "have nots" in the United States and who faced discrimination gradually attained an equal economic status to whites whose ancestry was Anglo-Saxon Protestant and who were the "haves". Such an evolution is consistent with a neoclassical view of the workings of competitive markets, assuming that the productive capacities of the different ethnic groups are equal and that the economy is sufficiently competitive.

A more specific application of economic principles to an analysis of discrimination involves Hispanic Americans, who are mainly recent immigrants. Their lower relative earnings may be rationalized by a theory of the determinants of earnings that assigns important roles to information about the labor market, to facility in the English language, and to education, measured by years of schooling. Such theories are qualitatively supported by empirical evidence. Whether the evidence shows that the quantitative gap in earnings between Hispanic and non-Hispanic whites is explained by these theories is not clear to me.

The difference in market earnings between men and women can be rationalized by economic theories of the gains from specialization and investment in human capital, combined with an assumption of voluntary choices by women to specialize in the home sector. This earnings gap, particularly between white men and women, is one of the largest and most time-persistent of the comparisons discussed in this paper (see Table 13.5).

In another paper [Cain (1985)] I have suggested that the theory of voluntary choice regarding labor market activities should lead to equality in *total incomes*

³² Hispanic immigrants are considered to be recent immigrants, and most blacks who came to the United States were not voluntary immigrants, so these two groups are not included in the group referred to as the earlier immigrants.

received by men and women, if not to equality in *labor market wages*. I assume equality in women's and men's productive capacity, in the nonpecuniary aspects of their work, and in their leisure consumption. I then test for the equality in income received by assuming that husbands and wives share their household income equally while married. Even with this assumption women were found to have a substantially lower present value of lifetime income than men: the ratios were between 0.7 and 0.9 (depending on various assumptions). These are, however, closer to unity than the usual measure of women-to-men ratios of wages, as reported in Table 13.6.

I also examined the total time spent in housework and market work combined for men and women. The data are weak, but the available evidence suggests near equality among husbands and wives [Cain (1984b)]. It is not clear how the inclusion of men and women without spouses present would affect this comparison. More women than men are likely to head single-parent families, and many of these women have the double burden of market work combined with a heavy workload at home, especially child care. On the other hand, those female heads-of-household who are recipients of public welfare tend not to work much in the market; indeed, the conditions of their welfare receipt discourage market work. Men who are single-parent heads-of-households are not likely to be on welfare. Another important unknown factor in my attempted comprehensive measure of economic well-being is the nonpecuniary utility (or disutility) that men and women obtain from their work.

The wages, earnings, and incomes of black workers and black households are substantially less than those of whites, and the conventional human capital variables, such as education, training, and health care leave much of the difference unexplained. Even if they explained more, the question would then be: Why is the market for such human capital investments functioning so poorly that blacks continue to be shortchanged? If whites find it profitable—in terms of higher earnings and better jobs—to make these investments, why are blacks' opportunities for these investments so curtailed? If the answer is not labor market discrimination, is it discrimination in the capital markets that supply funds or other sources of human capital investments? It is not scientifically satisfactory for economists to argue that labor market discrimination is minimal if they then have no explanation for how discrimination in capital markets creates and sustains the inequities we measure in the labor market.

The case of blacks in the United States appears to offer the strongest evidence for the reality of labor market discrimination and, given existing economic theories, for flaws in the competitive functioning of the market. In these respects, the case of blacks is at the other end of the spectrum from that of non-Hispanic early-immigrant groups. Economic discrimination, whether measured by average family incomes or by comparing wages when exogenous productivity factors have been held constant, is substantial for blacks and is nonexistent or insubstantial

for various former-immigrant white (and some Oriental) groups. For those groups, but not for blacks, the market has virtually eliminated the differences in economic attainment that were present decades ago.

5.2. The effects of discrimination on total output

The foregoing remarks refer to descriptive statistics, including regression analyses of earnings functions, and to the normative issue of equity. Let us turn to positive economics and the tasks of analysis, prediction, and explanation. One issue that has not been much studied is the implications of discrimination for economic efficiency as measured by total social income. I have elsewhere [Cain (1985)] addressed this question regarding discrimination against women, and my conclusions were embarrassingly thin. The neoclassical economist's convention, and perhaps it is an obligation, to take tastes – individual preferences – as given prevents the automatic translation of “different prices (wages) for the same good (labor)” into a loss in total social income (or total utility).³³ In a competitive economy in which tastes are the fundamental cause of discrimination there is no presumptive case for inefficiency.

Perhaps the underlying atomistic competitive model with only private (internalized) benefits and costs is too narrow. Alexis (1973, p. 297) distinguishes between discrimination, in which the discriminator is indifferent “to the welfare of the avoided [black] person”, and racism, where “the decision maker is not indifferent to the relative economic status of nonpreferred persons”. Alexis’s point may be extended to posit racist discrimination as a public “good” among the majority group, and this could explain the persistence of labor market discrimination. However, it is difficult to reconcile the idea of racism as a public good with the proliferation during the last 30 years of antidiscriminatory legislation and court decisions, which ought to reflect the public’s externalities regarding discrimination. On the other hand, there have also been numerous references in recent American politics to the “silent majority” among whites, who oppose the pro-civil-rights legislation. The purely political aspects of this topic are well beyond the scope of this chapter. The issue of externalities is not, but I do not pursue it.

Discrimination that results from private and government monopolies deserves further study, particularly if the term monopoly is extended to include collusive action that deprives minorities of access to various opportunities, some of which, like housing, may be only indirectly related to labor markets. Akerlof (1976)

³³Ambiguities about total welfare when tastes for discrimination are part of a person’s utility function are discussed by Thurow (1969, pp. 116–138), Toikka (1976), and Lundahl and Wadensjö (1984, pp. 81–108).

analyzes several models illustrating collusion, including conspiratorial acts and intimidation, that are sources of the oppression of minorities. Even if I am correct in my judgment that monopoly is not the predominant source of discriminatory wage differences, this does not imply that the benefits from attacking this source are less than the associated costs.

Aside from monopoly, the standard cases of market failure that point to inefficiencies that might be overcome through government intervention do not emerge from economic theories of discrimination. Perhaps an exception is the externalities of information concerning the productive capacities of minorities. Tests, licenses, certifications, and other such signals are used extensively in labor markets, and the private costs of obtaining accurate information about workers' productivity may be high relative to the private benefits, which are not necessarily appropriable as a private good. Clearly, those who believe in the equality of productive capacities across the groups under study will be more likely to believe in the benefits of more scientific information about productive capacities. The history of the stereotypes of inferiority among ethnic minorities in the United States is too familiar to cite, and the demise of these stereotypes regarding the earlier immigrant groups is part of the latter's success story.

In the main, however, I was not able to extract efficiency losses from the economic theories of discrimination [Cain (1985)], which reflects my agnostic view of these theories expressed in Section 3. Something is amiss. Discrimination in its many forms, not only economic, is widely believed to suppress the achievements of the minority group with no fully offsetting gains to the majority group. The economists that I know agree on this, yet conventional economic theories do not, to my knowledge, explain or analyze this widely shared belief.

Economists have prescribed limits for themselves in many policy spheres. Economics does not distinguish among the ethical merits of different tastes; between, say preferences for physical attractiveness or for race. As economists we have nothing to say about the justness of laws that prohibit an employer from refusing to hire someone on the basis of color but that permit hiring on the basis of physical attractiveness. As citizens we may, of course, have strong opinions about such matters.

5.3. Measuring the impact of policies

A more promising role for economic analysis lies in the measurements and methods that permit prediction. Empirical regularities, such as time trends, may be established and be useful even in the absence of fully developed theories. At a minimum, the measurements provide valuable data for monitoring progress or regress regarding discrimination.

A more ambitious form of prediction is that concerning policy instruments. The policies available to government may be classified into three categories: (a) macroeconomic (mainly monetary and fiscal policies affecting aggregate demand), (b) income transfers, and (c) microeconomic structural labor market policies. I discuss only the third. Supply-side structural policies typically comprise education and training programs. They have tended to be directed to low-income workers, with no special targeting to workers of a particular race or gender. In this regard, the supply-side policies tend to differ from the demand-side policies. Microeconomic demand-side policies might also be general, such as public employment programs or wage subsidies for unemployed and low-income workers, but the demand policies that have received most attention are those that directly forbid discrimination or promote preferential treatment of minority workers in hiring, placement, pay, or employment security. Preferential treatment is also called affirmative action.

Research aimed at evaluating these policies is abundant and controversial. See, for example, the proceedings of a conference on the labor market effects of antidiscrimination legislation that appeared in the *Industrial and Labor Relations Review* (vol. 29, July 1976). The essential difficulty in evaluating these programs is the classic problem of trying to make inferences from an uncontrolled experiment. We observe an outcome for a group of workers, some of whom participated in the program or, alternatively, had the program imposed on them. To establish causality between program status and the outcome, the factors that selected the workers into the program must be either (a) known and controlled for in the evaluation, or (b) known to be unrelated to the outcome.

It is difficult to know enough about the selection process and about all the causes of the outcome to justify either (a) or (b). Random assignment would satisfy condition (b), but this selection procedure is rare. Legislators and courts, therefore, seldom rely on the research of economists to determine the fate of government programs.

Decades of empirical research on labor markets, much of it like the research reported in Section 4, can be helpful in estimating the effects of a variety of policy variables on the earnings of workers, even if the research does not provide conclusive answers. Customarily we use cross-sectional data for empirical work, although the policy question is invariably one of predicting a change over time.

The implications for policy from cross-sectional research arose in the previous section. Some variables may be only minimally affected by policy, like "years since immigration" among the existing stock of immigrants; some partly affected, like "years of schooling"; and some almost wholly affected, like "participation in a government training program". Policies related to the cross-sectional findings will not be discussed, but some of the issues about how the above-mentioned variables affect outcomes in the labor market appear in time-series analyses of discrimination, which is the final topic of this chapter.

5.4. Results of time-series analyses

For my purposes the essential facts from time-series data that pertain to economic discrimination have been revealed by Table 13.5 in Section 2. This table shows two major challenges. One is the near-constant ratio of women's-to-men's wages over a 40-year period, using the data on earnings of year-round, full-time workers. The second is the slow increase in the ratio of black-to-white wages among men, which in 1982 was only 0.72. These trends in wages could be usefully supplemented with an analysis of trends in other measures of attainment in the labor market, including occupational attainment, labor force participation, and employment/unemployment rates, but space limitations preclude more than brief remarks.

The sharp increase in labor force participation rates (LFPRs) by women and the moderate decrease in LFPRs of men during the last 40 years have brought men and women into closer equality with respect to the quantity of time spent in market work, although men still spend about twice as many hours of their adult life in market work as do women [Cain (1984a)].

The increase in LFPRs for women has been the result of two trends which, as noted in Section 3, have contrasting effects on the trend in average wages of women: (a) women who work are working more continuously and for more years of their adulthood; (b) a larger fraction of women are entering and reentering the work force. Trend (a) should increase the average wage, because the wage should increase with experience and seniority. Trend (b) probably decreases the average wage because the composition of workers is altered by the influx of women with less-than-average experience – referred to as “adverse selection”. The “adverse selection” hypothesis is strongly advocated by Smith and Ward (1984), but see Fuchs (1984) for counter arguments and Mallan (1982) for counter evidence.³⁴

The LFPRs of men between the ages of 50 and 65 have declined during the last 20 years or so. Are these early retirements and disability-related retirements concentrated among low-wage workers? The substitution effect of wages on labor supply suggests a yes answer, but the income effect suggests otherwise. Retirement may be considered a “luxury good” that is selected by workers with above-average incomes. The net result of these contrasting effects on the trends among men needs to be studied.

The changing composition of the male labor force has also been examined in analyzing trends in wages of men, black and white. Butler and Heckman (1977)

³⁴ Further counter evidence to Smith and Ward is Maloney's finding that the wages of husbands and wives who work in every year covered by the Michigan Panel Study of Income Dynamics show declining women-to-men ratios from 1975 through 1980 [Maloney (1983, pp. 135–139)]. These data may, of course, simply reflect the contrasting age/earnings profiles shown in Figure 13.2. Whether the declining ratios imply labor market discrimination depends on whether they reflect voluntary choices of these wives to commit less effort than their husbands to the market sector.

Table 13.9
Median money incomes and income ratios for black and white male workers, 1948–1982,
in constant 1982 dollars

Year	White median\$	Black median\$	Ratio B/W	Year	White median\$	Black median\$	Ratio B/W
1948	\$10064	\$5465	0.54	1965	\$16185	\$8710	0.54
1949	10006	4844	0.48	1966	16631	9212	0.55
1950	10862	5899	0.54	1967	16901	9653	0.57
1951	11524	6346	0.55	1968	17388	10551	0.61
1952	11837	6487	0.55	1969	17812	10508	0.59
1953	12237	6760	0.55	1970	17428	10490	0.60
1954	12080	6011	0.50	1971	17248	10351	0.60
1955	12776	6724	0.53	1972	18029	11100	0.62
1956	13558	7113	0.52	1973	18360	11551	0.63
1957	13402	7096	0.53	1974	17330	11135	0.64
1958	13275	6614	0.50	1975	16679	10511	0.63
1959	13937	6561	0.47	1976	16849	10540	0.63
1960	14003	7367	0.53	1977	16889	10326	0.61
1961	14290	7385	0.52	1978	16945	10796	0.64
1962	14859	7318	0.49	1979	16363	10604	0.65
1963	15151	7874	0.52	1980	15612	9786	0.63
1964	15361	8708	0.57	1981	15172	9624	0.63
				1982	14748	9493	0.64

Source: U.S. Bureau of the Census, Current Population Reports, Series P-60, No. 142, *Money Income of Households, Families, and Persons in the United States: 1982* (Washington, D.C.: U.S. GPO, 1984), Table 40.

suggested that part of the rise in the ratio of black-to-white wages through the 1960s and into the early 1970s could be attributable to a selection of higher-earning workers among blacks, relative to whites. Their argument is as follows. Black men's LFPRs have declined more than have the LFPRs of white men. Assume that the men who drop out of the labor force tend to be low earners. The black male labor force would then have relatively fewer of the low earners remaining, and the average earnings of blacks – which is measured only for those in the labor force – will rise relative to that of whites.

The issue is not resolved, and it illustrates themes that I wish to stress in this survey: the need for closer attention to descriptive economic statistics about the labor market statuses of majority and minority groups, and the need to specify the purposes of one's analysis. A problem in charting trends is that the use of broad population groups may introduce exogenous compositional changes (like the age distribution) that should be held constant, while narrowing the comparison to various subgroups may reflect selection according to an endogenous characteristic (like full-time work status) that should not be held constant.

Table 13.9 shows another version of the black-to-white comparison among men, but this time for all men who worked, rather than for year-round, full-time workers, as in Table 13.5. Table 13.9 shows the time series for money incomes in

constant 1982 dollars for each year, 1948–1982. (The trend in money income is very similar to the trend in money earnings.) The dollar figures show the striking reversal from 1973 on, of the long-term growth in real incomes for both groups that prevailed from 1948 to 1973.

The overall picture regarding the black-to-white (B/W) ratios in Table 13.9 is similar to that summarized in Table 13.5, but the year-by-year statistics bring out more clearly the three periods of stability and change in the B/W ratios: 1948–1965, stability; 1966–1974, growth; 1975–1982, stability. Presumably a theory or an empirical evaluation of specific hypotheses about labor market discrimination against blacks should be able to explain these stylized facts.

Freeman (1981) discusses three main contending explanations, in addition to the Butler–Heckman hypothesis about selectivity in the composition of the populations. One is that the B/W ratio is pro-cyclical—rising during periods of prosperity, when unemployment is low, and falling during recessions. This is consistent with its growth in 1966–1974, when labor markets were relatively tight, and with the ratios that are low relative to surrounding years in the recession troughs of 1949, 1954, and 1958. However, the hypothesis is not supported by the stability of the ratios throughout the period of 1948–1965, when unemployment rates were relatively low, or by the behavior of the ratios in the later recession troughs, 1961, 1971, 1975, and 1982.

A second hypothesis is that the B/W ratios were affected by the surge of legal measures that may be said to have begun with an Executive Order in 1961 (No. 10925) that reinforced a somewhat dormant ban on discrimination by firms doing business with the federal government, followed by the Civil Rights Act of 1964 and subsequent legislation. Measuring the impact of these various forms of government intervention is difficult, however. How does one quantify the resources devoted to the intervention? How do we separate the effects of the legislation from the effects of the political and social climates that fostered the legislation? It is no surprise that the attempted evaluations of the legislation have not been conclusive.

A third hypothesis to explain the time series of black-to-white earnings ratios focuses on education, where this may be interpreted narrowly as years of schooling completed or broadly as a general indicator of human capital, including qualitative aspects of schooling as well as the training, information, and mobility that are affected by schooling. In either case, the emphasis is on the supply side of the labor market and the relative increases in the human capital of black men.

The role of education in this stream of research has had a curious history. Early quantitative studies based on the 1940, 1950, and 1960 censuses consistently showed two rather pessimistic regularities about male workers: (a) holding age constant, the B/W earnings ratio generally declined as years-of-schooling increased; (b) holding years of schooling constant, the B/W ratio generally declined as age increased. See Zeman (1955) as quoted in Becker, (1971,

Table 13.10
Median years of schooling completed and schooling ratios,
white and black men, various populations, 1940-1980.

Year	Population, by age and labor force status	Median years of educational attainment		
		White	Black	Ratio B/W
1940	All males 25+	8.6	5.7	0.66
	Males, 25-59	10.3	7.0	0.68
1950	All males 25+	9.3	6.8	0.73
	Males 25-29	12.0	8.6	0.72
1952	Males, 18+ in civilian labor force	10.8	7.2	0.67
	Males, 18+ in civilian labor force	11.9	8.3	0.70
1959	All males 25+	12.2	9.9	0.81
	Males, 25-29	12.6	12.2	0.97
	Males, 18+ in civilian labor force	12.1	11.1	0.89
1970	All males 25+	12.5	12.0	0.96
	Males, 25-29	12.9	12.6	0.98
	Males, 18+ in civilian labor force	12.7	12.2	0.96
1980				

Sources: "All males" and "Males, 25-29"; U.S. Bureau of the Census, *Statistical Abstract of the United States: 1982-83*, 103rd edition (Washington, D.C.: U.S. GPO, 1983) p. 143; "Males, 18+ in civilian labor force", U.S. Department of Labor, *Handbook of Labor Statistics 1978*, Bulletin 2000 (Washington, D.C.: U.S. GPO, 1979) p. 68; "Males, 16+ in civilian labor force", U.S. Department of Labor, Bureau of Labor Statistics, *Handbook of Labor Statistics*, Bulletin 2175 (Washington, D.C.: U.S. GPO, 1981), pp. 152-153.

p. 111) for 1940, Anderson (1955) and Freeman (1973b, p. 85) for 1950, and Hanoch (1967) for 1960.

At face value, (a) implies that an equal growth in educational attainment over time would widen the B/W difference in earnings. However, the increase in educational attainment by black men has exceeded that of white men, particularly between 1960 and 1980 (see Table 13.10). This period includes, but does not coincide with, the period of the rise in the earnings ratios, 1967-1974. By 1980 the B/W differences in median educational attainments had been virtually eliminated, although there may well remain differences in the quality of schooling.

In addition to the relative increase in the quantity of schooling obtained by blacks, a new development in research findings in the 1970s re-emphasized the importance of schooling. Recent studies show that the wage returns to schooling were becoming more equalized between blacks and whites, although this was mainly true for the young age groups and those with some college [see Smith and Welch (1977), Freeman (1977)]. Like the vintage hypothesis regarding age effects, these new and higher education effects for black men relative to white men remain to be tested in the years to come.

This brief survey of proposed explanations for the trend in black-to-white male earnings illustrates the tentativeness of empirical regularities regarding labor market discrimination and the consequent difficulty in drawing policy implications. Estimated relations from cross-sections at different times show widely varying results, and the time series, with relatively few observations and many competing hypotheses, do not yield firm empirical regularities either.

5.5. Final word

The economics of discrimination is a particularly complex subject. My judgment is that the theories of discrimination have been useful for providing definitions and for suggesting measurements of discrimination but not for providing convincing explanations of the phenomenon nor of its patterns. The econometric work has also been useful, but to my eyes more so for its descriptive content than for testing hypotheses or for providing estimates of causal relations.

References

- Abowd, John M. and Mark Killingsworth (1982) "Employment, wages, and earnings of Hispanics in the federal and non-federal sectors: methodological issues and their empirical consequences", in: G. Borjas and M. Tienda, eds., *Hispanic labor conference*. Washington, D.C.: National Commission on Employment Policy. (Subsequently published in G. J. Borjas and M. Tienda, eds., *Hispanics in the U.S. Economy*. Orlando, Fla.: Academic Press, (1985, 77-125.)
- Aigner, Dennis J. and Glen G. Cain (1977) "Statistical theories of discrimination in the labor market", *Industrial and Labor Relations Review*, 30:175-187.
- Akerlof, George A. (1976) "The economics of caste and of the rat race and other woeful tales", *Quarterly Journal of Economics*, 94:599-617.
- Alchian, Armen A. and Reuben Kessel (1962) "Competition, monopoly and the pursuit of pecuniary gain", in: *Aspects of labor economics*, Conference Volume of the Universities-National Bureau Committee for Economic Research. Princeton, N.J.: Princeton University Press, 156-175.
- Alexis, Marcus (1973) "A theory of labor market discrimination with interdependent utilities", *American Economic Review*, 63:296-302
- Anastasi, Anne (1969) "Differential psychology", in: *Encyclopedia Britannica*. Chicago: Encyclopedia Britannica, Inc., William Benton, vol. 7.
- Anderson, C. Arnold (1955) "Regional and racial differences in the relation between income and education", *The school review*, 63:38-45.
- Arrow, Kenneth (1972) "Models of job discrimination", and "Some mathematical models of race in the labor market", in: A. H. Pascal, ed., *Racial discrimination in economic life*. Lexington, Mass.: Lexington Books, 83-102 and 187-204.

- Arrow, Kenneth (1973) "The theory of discrimination", in: O. A. Ashenfelter and A. Rees, eds., *Discrimination in labor markets*. Princeton, N.J.: Princeton University Press, 3–33.
- Ashenfelter, Orley (1972) "Racial discrimination and trade unions", *Journal of Political Economy*, 80:435–464.
- Ashenfelter, Orley and Timothy Hannan (1986) "Sex discrimination and product market competition: the case of the banking industry", *Quarterly Journal of Economics*, 101: 149–173.
- Astin, Helen S. and Alan E. Bayer (1972) "Sex discrimination in academe", *Educational Record*, 53:101–118.
- Baldus, D. C. and J. W. L. Cole (1980) *Statistical proof of discrimination*. New York: McGraw Hill.
- Becker, Gary S. (1971) *The economics of discrimination*. Chicago: The University of Chicago Press. (Original edition, 1957.)
- Becker, Gary S. (1959) "Union restrictions on entry", in: P. D. Bradley, ed., *The public stake in union power*. Charlottesville, Va.: University of Virginia Press, 209–224.
- Becker, Gary S. (1968) "Discrimination, economic", in: D. L. Sills, ed., *International encyclopedia of the social sciences*. New York: Macmillan and Free Press, vol. 4, 208–210. (Reprint edition, 1972.)
- Bergmann, Barbara R. (1974) "Occupational segregation, wages and profits when employers discriminate by race or sex", *Eastern Economic Journal*, 1:103–110.
- Blau, Francine D. (1977) *Equal pay in the office*. Lexington, Mass.: D.C. Heath and Co.
- Blinder, Alan S. (1973) "Wage discrimination: reduced form and structural variables", *Journal of Human Resources*, 8:436–455.
- Bunting, Robert L. (1962) *Employer concentration in local labor markets*. Chapel Hill: University of North Carolina Press.
- Butler, Richard and James J. Heckman (1977) "Government's impact on the labor market status of black Americans: a critical review", in: *Equal rights and industrial relations*. Madison, Wis.: University of Wisconsin, Industrial Relations Research Association, 235–281.
- Cain, Glen G. (1976) "The challenge of segmented labor market theories to orthodox theory: a survey", *Journal of Economic Literature*, 14:1215–1257.
- Cain, Glen G. (1984a) "Lifetime measures of labor supply of men and women", Discussion Paper No. 749-84, Institute for Research on Poverty, University of Wisconsin-Madison.
- Cain, Glen G. (1984b) "Women and work: trends in time spent in housework", Discussion Paper No. 745-84, Institute for Research on Poverty, University of Wisconsin-Madison.
- Cain, Glen G. (1985) "Welfare economics of policies toward women", *Journal of Labor Economics*, 3, part II:375–396.
- Chiswick, Barry R. (1973) "Racial discrimination in the labor market: a test of alternative hypotheses", *Journal of Political Economy*, 81:1330–1352.
- Chiswick, Barry R. (1974) *Income inequality*. New York: National Bureau of Economic Research, distributed by Columbia University Press.
- Chiswick, Barry R. (1983) "The earnings and human capital of American Jews", *Journal of Human Resources*, 18:313–336.
- Cohen, Malcolm S. (1971) "Sex differences in compensation", *Journal of Human Resources*, 6:434–447.
- Comanor, William S. (1973) "Racial discrimination in American industry", *Economica*, 40:363–378.
- Conway, Delores A. and Harry V. Roberts (1983) "Reverse regression, fairness and employment discrimination", *Journal of Business and Economic Statistics*, 1:75–85.
- Corcoran, Mary and Greg J. Duncan (1979) "Work history, labor force attachment, and earnings differences between the races and sexes", *Journal of Human Resources*, 14:497–520.
- Duncan, Otis Dudley (1968) "Inheritance of poverty or inheritance of race?", in: D. P. Moynihan, ed., *On understanding poverty*. New York: Basic Books, 85–110.
- Featherman, David L. and Robert M. Hauser (1976) "Sexual inequalities and socioeconomic achievement in the U.S., 1962–1973", *American Sociological Review*, 41:462–483.
- Finkelstein, Michael O. (1980) "The judicial reception of multiple regression studies in race and sex discrimination cases", *Columbia Law Review*, 80:737–754.
- Fisher, Franklin M. (1980) "Multiple regression in legal proceedings", *Columbia Law Review*, 80:702–736.
- Flanagan, Robert J. (1973) "Racial wage discrimination and employment segregation", *Journal of Human Resources*, 8:465–471.
- Flanagan, Robert J. (1974) "Labor force experience, job turnover, and racial wage differentials", *Review of Economics and Statistics*, 56:521–529.

- Freeman, Richard B. (1973a) "Decline of labor market discrimination and economic analysis", *American Economic Review*, 63:280–286.
- Freeman, Richard B. (1973b) "Comment", in: O. Ashenfelter and A. Rees, eds., *Discrimination in labor markets*. Princeton, N.J.: Princeton University Press, 82–87.
- Freeman, Richard B. (1977) *Black elite: the new market for highly qualified black Americans*. New York: McGraw-Hill.
- Freeman, Richard B. (1981) "Black progress after 1964: who has gained and why?", in: S. Rosen, ed., *Studies in labor markets*. Chicago: University of Chicago Press, 247–294.
- Fuchs, Victor R. (1971) "Differences in hourly earnings between men and women", *Monthly Labor Review*, 94:9–15.
- Fuchs, Victor R. (1984) "His and hers: gender differences in work and income, 1959–1979", Working Paper No. 1501, National Bureau of Economic Research, Cambridge, Mass. Forthcoming in *Journal of Labor Economics*.
- Goldberg, Matthew S. (1982) "Discrimination, nepotism, and long-run wage differentials", *Quarterly Journal of Economics*, 96:307–319.
- Goldberger, Arthur S. (1984) "Reverse regression and salary discrimination", *Journal of Human Resources*, 19:293–318.
- Gould, William B. (1977) *Black workers in white unions: job discrimination in the United States*. Ithaca, N.Y.: Cornell University Press.
- Greeley, Andrew M. (1976) *Ethnicity, denomination and inequality*. Beverly Hills, Calif.: Sage Publications.
- Greeley, Andrew M. (1981) *The Irish Americans*. New York: Harper and Row.
- Green, Carol A. and Marianne A. Ferber (1984) "Employment discrimination: an empirical test of forward versus reverse regression", *Journal of Human Resources*, 19:557–569.
- Grenier, Gilles (1984) "The effect of language characteristics on the wage of Hispanic-American males", *Journal of Human Resources*, 19:35–52.
- Gwartney, James D. and Richard Stroup (1973) "Measurement of employment discrimination according to sex", *Southern Economic Journal*, 39:575–587.
- Hanoch, Gloria (1967) "An economic analysis of earnings and schooling", *Journal of Human Resources*, 2:310–329.
- Hill, Herbert (1977) *Black labor and the American legal system*. Washington, D.C.: The Bureau of National Affairs.
- Hoffman, Saul (1979) "Black-white life cycle earnings differences and the vintage hypothesis: a longitudinal analysis", *American Economic Review*, 69:855–867.
- Johnson, George E. (1975) "Economic analysis of trade unionism", *American Economic Review, Papers and Proceedings*, 65:23–28.
- Johnson, George E. and Frank P. Stafford (1974) "The earnings and promotion of women faculty", *American Economic Review*, 64:888–903.
- Jones, F. L. (1983) "On decomposing the wage gap: a critical comment on Blinder's method", *Journal of Human Resources*, 18:126–130.
- Kessel, Reuben A. (1958) "Price discrimination in medicine", *Journal of Law and Economics*, 1:20–53.
- Kohen, Andrew and Roger Roderick (1975) "The effect of race and sex discrimination on early career earnings", unpublished paper, Center for Human Resource Research, Ohio State University, Columbus, Ohio.
- Landes, William (1968) "The economics of fair employment laws", *Journal of Political Economy*, 76:507–552.
- Lewis, H. Gregg (1959) "Competitive and monopoly unionism", in: P. D. Bradley, ed., *The public stake in union power*. Charlottesville, Va.: University of Virginia Press, 181–208.
- Lucas, Robert E. B. (1974) "The distribution of job characteristics", *Review of Economics and Statistics*, 56:530–540.
- Luksetich, W. A. (1979) "Market power and sex discrimination in white-collar employment", *Review of Social Economy*, 37:211–224.
- Lundahl, Mats and Eskil Wadensjö (1984) *Unequal treatment: a study in the neo-classical theory of discrimination*. London: Croom Helm.
- Lundberg, Shelly J. and Richard Startz (1983) "Private discrimination and social intervention in competitive labor markets", *American Economic Review*, 73:340–347.
- Madden, Janice F. (1973) *The economics of sex discrimination*. Lexington, Mass.: D.C. Heath and Co.

- Malkiel, Burton G. and Judith A. Malkiel (1973) "Male-female pay differentials in professional employment", *American Economic Review*, 63:693–705.
- Mallan, Lucy B. (1982) "Labor force participation, work experience, and the pay gap between men and women", *Journal of Human Resources*, 17:437–448.
- Maloney, Timothy J. (1983) "The cyclical labor supply response of married women and disequilibrium unemployment", unpublished Ph.D. dissertation, Department of Economics, University of Wisconsin-Madison.
- Marshall, F. Ray (1965) *The Negro and organized labor*. New York: John Wiley.
- Marshall, F. Ray (1974) "The economics of racial discrimination: a survey", *Journal of Economic Literature*, 12:849–871.
- Masters, Stanley H. (1975) *Black–white income differentials*. New York: Academic Press.
- McCall, John J. (1972) "The simple mathematics of information, job search, and prejudice", in: A. Pascal, ed., *Racial discrimination in economic life*. Lexington, Mass.: Lexington Books, 205–224.
- Mincer, Jacob and Solomon Polachek (1974) "Family investments in human capital earnings of women", *Journal of Political Economy*, 82, part 2:S76–S108.
- Myrdal, Gunnar (1944) *An American dilemma*. New York: Harpers.
- Northrup, Herbert R. (1944) *Organized labor and the Negro*. New York: Harpers.
- Oaxaca, Ronald L. (1973) "Male-female wage differentials in urban labor markets", *International Economic Review*, 14:693–709.
- O'Neill, June (1983a) "The determinants and wage effects of occupational segregation", working paper, Urban Institute, Washington, D.C.
- O'Neill, June (1983b) "The trend in sex differentials in wages", paper presented at the Conference on Trends in Women's Work, Education, and Family Building, Urban Institute, Washington, D.C.
- O'Neill, June (1984) "Earnings differentials: empirical evidence and causes", in: G. Schmid and R. Weitzel, eds., *Sex discrimination and equal opportunity*. London: Gower Publishing Co., 69–91.
- Oster, Sylvia M. (1975) "Industry differences in discrimination against women", *Quarterly Journal of Economics*, 89:215–229.
- Petersen, William (1978) "Chinese Americans and Japanese Americans", in: T. Sowell, ed., *Essays and data on American ethnic groups*. Washington, D.C.: The Urban Institute, 65–106.
- Phelps, Edmund S. (1972) "The statistical theory of racism and sexism", *American Economic Review*, 62:659–661.
- Piore, Michael J. (1970) "Jobs and training", in: S. H. Beer, ed., *The State and the Poor*. Cambridge, Mass.: Winthrop Press, 53–83.
- Polachek, Solomon W. (1979) "Occupational segregation among women: theory, evidence, and prognosis", in: C. B. Lloyd, E. S. Andrews and C. L. Gilroy, eds., *Women in the Labor Market*. New York: Columbia University Press, 137–157.
- Reich, Michael (1971) "The economics of racism", in: David M. Gordon, ed., *Problems in economy: an urban perspective*. Lexington, Mass.: D.C. Heath and Co., 107–113.
- Reich, Michael (1981) *Racial inequality*. Princeton, N.J.: Princeton University Press.
- Reimers, Cordelia W. (1983) "Labor market discrimination against hispanic and black men", *Review of Economics and Statistics*, 65:570–579.
- Riley, John G. (1975) "Competitive signaling", *Journal of Economic Theory*, 10:174–186.
- Roberts, Harry V. (1980) "Statistical biases in the measurement of employment discrimination", in: E. R. Livernash, eds., *Comparable worth: issues and alternatives*, Washington, D.C.: Equal Employment Advisory Council, 173–195.
- Robinson, Joan (1934) *The economics of imperfect competition*. London: Macmillan.
- Roos, Patricia A. (1981) "Sex stratification in the workplace: male-female differences in economic returns to occupations", *Social Science Research*, 10:195–224.
- Ross, Malcolm (1948) *All manner of men*. New York: Reynal and Hitchcock.
- Rothschild, Michael and Joseph E. Stiglitz (1982) "A model of employment outcomes illustrating the effect of the structure of information on the level and the distribution of income", *Economics Letters*, 10:231–236.
- Sanborn, Henry (1964) "Pay differences between men and women", *Industrial and Labor Relations Review*, 17:534–550.
- Sawhill, Isabel (1973) "The economics of discrimination against women: some new findings", *Journal of Human Resources*, 8:383–396.

- Smith, James P. and Michael P. Ward (1984) "Women's wages and work in the twentieth century", Report R-3119-NICHD, RAND, Santa Monica, Calif.
- Smith, James P. and Finis Welch (1977) "Black-white male wage ratios: 1960-1970", *American Economic Review*, 67:323-338.
- Smith, James P. and Finis Welch (1978) "Race differences in earnings: a survey and new evidence", unpublished ms., R-2295-NSF, RAND, Santa Monica, Calif.
- Sowell, Thomas (1981) *Ethnic Americans*. New York: Basic Books.
- Spence, A. Michael (1973) "Job market signaling", *Quarterly Journal of Economics*, 87:355-374.
- Stiglitz, Joseph E. (1973) "Approaches to the economics of discrimination", *American Economic Review, Papers and Proceedings*, 63:287-295.
- Stiglitz, Joseph E. (1974) "Theories of discrimination and economic policy", in: G. M. von Furstenberg, A. R. Horowitz, and B. Harrison, eds., *Patterns of racial discrimination. Volume B: employment and income*. Lexington, Mass.: D.C. Heath and Co., 5-26.
- Suter, Larry E. and Herman P. Miller (1973) "Income differences between men and career women", *American Journal of Sociology*, 78:962-974.
- Thurow, Lester C. (1969) *Poverty and discrimination*. Washington, D.C.: The Brookings Institution.
- Thurow, Lester C. (1975) *Generating inequality*. New York: Basic Books.
- Toikka, Richard S. (1976) "The welfare implications of Becker's discrimination coefficient", *Journal of Economic Theory*, 13:472-477.
- Tomes, Nigel (1984) "The effects of religion and denomination on earnings and the returns to human capital", *Journal of Human Resources*, 19:472-488.
- Treiman, Donald J. and Heidi I. Hartmann, eds. (1981) *Women, work, and wages: equal pay for jobs of equal value*. Washington, D.C.: National Academy Press.
- Treiman, Donald J. and Kermit Terrell (1975) "Sex and the process of status attainment: a comparison of working women and men", *American Sociological Review*, 40:174-200.
- U.S. Bureau of the Census (1980) *Statistical abstract of the United States, 1980*. Washington, D.C.: U.S. GPO, 101st ed.
- U.S. Bureau of the Census (1983a) Current population reports, Series P-20, No. 372, *Marital status and living arrangements, March 1981*. Washington, D.C.: U.S. GPO.
- U.S. Bureau of the Census (1983b) "Current population reports", Series P-60, No. 137, *Money income of households, families, and persons in the United States: 1981*. Washington, D.C.: U.S. GPO.
- U.S. Bureau of the Census (1983c) *Statistical abstract of the United States, 1984*. Washington, D.C.: U.S. GPO, 104th ed.
- U.S. Bureau of the Census (1984) "Current population reports", Series P-60, No. 142, *Money income of households, families, and persons in the United States: 1982*. Washington, D.C.: U.S. GPO.
- U.S. Department of Labor, Bureau of Labor Statistics (1979) *Earnings and other characteristics of organized labor, May 1977*, Report 556. Washington, D.C.: U.S. GPO.
- U.S. Department of Labor, Bureau of Labor Statistics (1982) *Labor force statistics derived from the current population survey: a databook, Volume I*, Bulletin 2096. Washington, D.C.: U.S. GPO.
- Welch, Finis (1973) "Education and racial discrimination", in: O. A. Ashenfelter and A. Rees, eds., *Discrimination in labor markets*. Princeton, N.J.: Princeton University Press, 43-81.
- Welch, Finis (1975) "Human capital theory: education, discrimination, and life cycles", *American Economic Review*, 65:63-73.
- Zabalza, Anthony and J. L. Arrufat (1983) "Wage differences between married men and married women in Great Britain: the depreciation effect of nonparticipation", unpublished Discussion Paper No. 151, Centre for Labour Economics, London School of Economics.
- Zeman, Mordechai (1955) "A quantitative analysis of white-nonwhite income differentials in the United States in 1939", unpublished Ph.D. dissertation, University of Chicago.