

Hellerstein, Judith K.; Neumark, David

**Working Paper**

## Using Matched Employer-Employee Data to Study Labor Market Discrimination

IZA Discussion Papers, No. 1555

**Provided in Cooperation with:**

IZA – Institute of Labor Economics

Suggested Citation: Hellerstein, Judith K.; Neumark, David (2005) : Using Matched Employer-Employee Data to Study Labor Market Discrimination, IZA Discussion Papers, No. 1555, Institute for the Study of Labor (IZA), Bonn

This Version is available at:

<http://hdl.handle.net/10419/21331>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

IZA DP No. 1555

## Using Matched Employer-Employee Data to Study Labor Market Discrimination

Judith Hellerstein  
David Neumark

April 2005

# Using Matched Employer-Employee Data to Study Labor Market Discrimination

**Judith Hellerstein**

*University of Maryland*

**David Neumark**

*Public Policy Institute of California  
and IZA Bonn*

Discussion Paper No. 1555  
April 2005

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0  
Fax: +49-228-3894-180  
Email: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### Using Matched Employer-Employee Data to Study Labor Market Discrimination\*

Wage gaps between individuals of different races, sexes, and ethnicities have been documented and replicated extensively, and have generated a long history in labor economics research of empirical tests for labor market discrimination. The most widely-used approach to test for labor market discrimination is based on wage regressions estimated at the level of individual workers, with the estimate of discrimination inferred from the residual race, sex, or ethnic group differential in wages that remains unexplained after including a wide array of proxies for productivity. What is absent from the residual wage approach – and in our view leaves the approach vulnerable to being regarded as uninformative regarding discrimination – is any directly observable measure of productivity with which to adjust differentials in wages in trying to infer whether a particular group suffers from discrimination. The ideal solution would be individual-level productivity data that can be compared with wages. Any of the variables that differ across groups and are unobserved in the residual wage regression approach should affect wages and productivity equally, and hence not bias the test. However, such data are extremely rare, in large part because individual productivity is often unobservable and seldom measured. This chapter focuses on the use of matched employer-employee data sets to carry out a version of this ideal test, but at the establishment level. When these data sets permit the measurement of the demographic characteristics of establishments' workforces, as well as the estimation of production functions, they can be used to infer productivity differentials between workers in different groups. Comparisons of these productivity differentials with wage differentials then provide versions of the ideal test for discrimination at the establishment level. In addition to providing tests of discrimination, matched employer-employee data sets have proven useful in studying other questions that arise in the economics of discrimination, including measuring labor market segregation and assessing its consequences, and examining hypotheses or predictions that are central to economic models of discrimination.

JEL Classification: J71

Keywords: labor market discrimination, matched employer-employee data

Corresponding author:

David Neumark  
Public Policy Institute of California  
500 Washington St., Suite 800  
San Francisco, CA 94111  
USA  
Email: [neumark@ppic.org](mailto:neumark@ppic.org)

---

\* We are grateful to Kimberly Bayard and Kenneth Troske for past collaboration, to Joel Elvery and Gigi Foster for research assistance, and to David Levine for helpful comments. Some of the research described in this paper was supported by the National Science Foundation (grant SBR95-10876) and by the Russell Sage Foundation.

## **I. Introduction**

Wage gaps between individuals of different races, sexes, and ethnicities have been documented and replicated extensively, and have generated a long history in labor economics research of empirical tests for labor market discrimination. The most widely-used approach to test for labor market discrimination is based on wage regressions estimated at the level of individual workers, with the estimate of discrimination inferred from the residual race, sex, or ethnic group differential in wages that remains unexplained after including a wide array of proxies for productivity. Such estimates are open to the criticism that the proxies do not adequately capture group-specific differences in productivity, leading to disputes over the interpretation of the estimates that cannot be resolved with the types of data typically used in this approach. What is absent from the residual wage approach—and in our view leaves the approach vulnerable to being regarded as uninformative regarding discrimination—is any directly observable measure of productivity with which to adjust differentials in wages in trying to infer whether a particular group suffers from discrimination. The ideal solution would be individual-level productivity data that can be compared with wages. Any of the variables that differ across groups and are unobserved in the residual wage regression approach should affect wages and productivity equally, and hence not bias the test. However, such data are extremely rare, in large part because individual productivity is often unobservable and seldom measured.

This chapter focuses on the use of matched employer-employee data sets to carry out a version of this ideal test, but at the establishment level. When these data sets permit the measurement of the demographic characteristics of establishments' workforces, as well as the estimation of production functions, they can be used to infer productivity differentials between workers in different groups. Comparisons of these productivity differentials with wage differentials then provide versions of the ideal test for discrimination at the establishment level.

In addition to providing tests of discrimination, matched employer-employee data sets have proven useful in studying other questions that arise in the economics of discrimination, including

measuring labor market segregation and assessing its consequences, and examining hypotheses or predictions that are central to economic models of discrimination.

The unifying theme of this chapter, then, is that matched employer-employee data sets hold out considerable promise for generating empirical evidence regarding labor market discrimination that cannot be gleaned from the individual-level data that has dominated most research on the topic. The chapter proceeds in a few steps. First, it explains some of the advantages that matched employer-employee data offer in studying discrimination. Second, it provides a guide to matched employer-employee data sets, focusing on the United States but also discussing data sets constructed for other countries. Third, it reviews some of the work that has been done using matched employer-employee data, again focusing primarily on the United States. Finally, it discusses new directions in which we think the construction and analysis of matched employer-employee data sets is likely to prove most useful in advancing our understanding of labor market discrimination.

## **II. The Wage Residual Approach and Other Methods**

### *Theoretical Motivation*

The wage residual approach to discrimination can be motivated by Becker's seminal model of employer discrimination (Becker, 1971). This is discussed in some detail because it is also at the core of the test for discrimination based on matched employer-employee data. This model also yields some insights regarding the persistence of discrimination over the longer run; these are discussed later. The assumption in the employer discrimination model is that employers dislike hiring a particular group, such as women. The implication is that when a woman is hired, an employer considers the cost to be both the wage and the disutility from hiring a woman. Thus, in the simplest case where men and women are perfect substitutes in production, employers can be thought of as maximizing a utility function of the form

$$U_D = U(\pi, F) = \pi - d \cdot F = f(M+F) - w_M M - w_F F - d \cdot F ,$$

where  $d$  is a constant  $> 0$ , reflecting discriminatory tastes against women, and the  $D$  subscript denotes a discriminating employer. The first-order conditions are

$$MP_L = w_M , MP_L = w_F + d .$$

Since  $d$  is positive, the only equilibrium in which men and women are employed is the one in which  $w_M = w_F + d$ , and if  $d = 0$  we must have  $w_M = w_F$ .

Thinking about the equilibrium requires some care. If  $d$  is the same for all employers, then when  $w_M = w_F + d$  employers are indifferent between hiring male and female labor, so we expect all employers to hire both men and women. In contrast, if, as seems more likely,  $d$  varies across employers (indexed by  $i$ ), then some equilibrium wage differential between men and women will be generated such that if  $d_i > w_M - w_F$ , employer  $i$  hires only men, and if  $d_i < w_M - w_F$ , employer  $i$  hires only women. Thus, one problem with Becker's employer discrimination model (as an explanation of the wage differential) is that, in the real world, segregation with almost all firms hiring only one sex or the other is generally not observed.

However, this is a result of the specific form of the utility function. If employers care about the relative number of female employees, rather than the absolute number, then the utility function is

$$U_D = f(M + F) - w_M M - w_F F - d \cdot (F/M) ,$$

in which case the first-order conditions are

$$MP_L = w_F + d/M , \quad MP_L = w_M - d \cdot F/M^2 .$$

Clearly  $w_M$  will exceed  $w_F$  in equilibrium. But now employers can effectively adjust the cost of hiring women by adjusting the relative number of women, so that for any equilibrium wage ratio, firms should be willing to hire some men and some women. Of course, those with smaller  $d$ 's will hire relatively more women. So versions of the employer discrimination model can explain the existence of wage differentials without requiring near-complete sex segregation of firms. In general, then, in employer discrimination models of this type, the preferences of employers impact hiring decisions, which in equilibrium generate a market wage differential.

### *Empirical Approach*

As the previous first-order conditions suggest, the wage discrimination that results in the employer discrimination model takes the form of unequal wages paid to men and women with the same marginal products. As shown explicitly in Neumark (1988), this model can be used to motivate directly

the wage residual approach to discrimination developed by Oaxaca (1973) and Blinder (1973). The simplest approach is to estimate a wage equation of the form

$$\ln(w) = X\beta + F\gamma + \varepsilon ,$$

where  $X$  is a vector of control variables presumed to capture individual differences in productivity, and  $F$  is a dummy variable equal to one for women. The implication of this specification is that the wage differential between men and women is constant (in percentage terms) for all values of the  $X$ 's. However, this may be unduly restrictive, failing to capture, for example, variation in the wage gap with the level of education or experience. Thus, the Oaxaca-Blinder procedure is to estimate two wage regressions

$$\ln(w) = X\beta_F + \varepsilon_F ,$$

and

$$\ln(w) = X\beta_M + \varepsilon_M ,$$

where  $X$  includes the unit vector.

In this case the decomposition of the wage differential varies depending on where it is evaluated, and the residual part of the wage differential now includes not just the difference between the intercepts, but also the difference between the coefficients on  $X$ . Since regression lines go through the means, if  $b_j, j = F, M$  are the vectors of estimates of  $\beta_F$  and  $\beta_M$ , and  $X_j'$  denotes the mean of  $X_j$ , then the difference between the average log wages of men and women can be written in one of two ways:

$$X_M' b_M - X_F' b_F = X_M' (b_M - b_F) + (X_M' - X_F') b_F ,$$

or

$$X_M' b_M - X_F' b_F = X_F' (b_M - b_F) + (X_M' - X_F') b_M .$$

Intuitively, what is the difference between these alternative decompositions? The last term in each equation captures the difference in wages that would remain even if the wage regressions were the same for men and women (i.e.,  $\beta_F = \beta_M$ ), because there would still be a difference in characteristics between men and women that might lead to a wage differential. Estimating what the contribution of this difference in characteristics would be requires an assumption about what the wage structure would be in the absence of discrimination. The first decomposition assumes that the wage structure in the absence of



discrimination would be  $\beta_F$ , whereas the second assumes it would be  $\beta_M$ . Conversely, using the first decomposition, for example,  $X_M' (b_M - b_F)$  measures the share of the wage gap attributable to discrimination. Whether or not this differs from zero is used to establish whether there is any evidence of discrimination.

### *Criticisms of the Wage Residual Approach*

The wage residual approach is subject to numerous criticisms. One important criticism is that it leads to ambiguous empirical answers because of an inherent arbitrariness in deciding the “base” with which to compare the present wage structure. To see this, note that a more general decomposition is

$$X_M' b_M - X_F' b_F = [X_M' (b_M - b) - X_F' (b_F - b)] + (X_M' - X_F') b ,$$

where  $b$  is the “no discrimination” wage structure. If  $b = b_F$ , then the first decomposition results, while the second results if  $b = b_M$  (see Neumark, 1988; Cotton, 1988). As evidence that this can be quantitatively important, in a study of pay discrimination in academia, in an admittedly extreme example Ferber and Green (1982) find that using  $b_M$  results in an estimate of 2% of the sex wage differential reflecting discrimination, while using  $b_F$  results in an estimate of 70%. In addition, as shown in Neumark (1988), there are alternative formulations of the assumption regarding employers’ discriminatory tastes that lead to different decompositions. The point of this is not that there is a “better” decomposition than those proposed by Oaxaca and Blinder, but rather that without some assumption—untested and likely arbitrary—regarding what the wage structure would look like in the absence of discrimination, these decompositions should perhaps not be interpreted as measuring discrimination, even ignoring the additional problems discussed below.<sup>1</sup>

While the foregoing criticism of the residual wage approach is technical, perhaps the most fundamental problem is that the control variables that are included in  $X$  may not fully capture marginal productivity differences. In particular, when there are unmeasured productivity differences that are systematically different across groups (e.g., race or sex), spurious evidence of discrimination can result.

---

<sup>1</sup>However, Oaxaca and Ransom (1994) offer some arguments for preferring Neumark’s (1988) decomposition, and provide an integrated discussion of the alternative decompositions.

Perhaps the best known example of such an argument is Becker's (1985) work on effort allocation, which explicitly posits an unobservable-effort-that affects wages and is supplied to the labor market with less intensity by women than by men.<sup>2</sup> There are other examples of such arguments in the literature. With respect to sex differences, Mincer and Polachek (1974) criticized early wage residual estimates using potential experience as a control variable because, given the more extensive labor market interruptions of women, potential experience overstates the actual experience-and hence human capital-of women relative to men. More recently, Light and Ureta (1995) documented the different patterns of accumulation of labor market experience for men's and women's wages, suggesting that even with an actual experience measure this issue remains important. With regard to race differences, Neal and Johnson (1996) and O'Neill (1990) argue that pre-market factors (as captured in test scores) account for much of the wage gap between blacks and whites.

The bottom line, in our view, is that because one can always tell a story about an unobservable that is related to productivity (although Becker's story may be a particularly good one), deciding whether residual wage estimates capture discrimination may be more an act of faith than an act of science. This suggests, in turn, that we should look to other methods.

#### *Audit Studies as a Response*

One response to the absence of convincing evidence from the wage residual approach has been audit studies, which are based on comparisons of outcomes (usually job interviews) for matched job applicants of different races or sexes (see, e.g., Turner, et al., 1991; Neumark, 1996).<sup>3</sup> In a sense, audit studies can be viewed as an alternative approach to the problem of missing data on productivity. Residual wage discrimination studies attempt to solve this problem by means of regression analysis, starting with

---

<sup>2</sup>As Darity and Mason (1998) point out, the omitted variables do not necessarily create a bias towards finding evidence of discrimination.

<sup>3</sup>Goldin and Rouse (2000) has many features in common with audit studies, although it is actually an unusual case of a natural experiment where the sex of a job applicant was, at times, not revealed to those making hiring decisions. Bertrand and Mullainathan (2004) perform what has been termed a "correspondence study" in the earlier literature, using resumes rather than actual visits to apply for jobs.

men and women or blacks and whites that are on average different across many dimensions, and trying to introduce enough variables such that, after controlling for these variables, the only remaining difference on average is race or sex. Of course, since residual wage discrimination studies use data drawn from a population of individuals in which real differences exist between groups, the need to adjust via regression is unavoidable. The audit study approach, in contrast, creates an artificial pool of labor market participants among whom there are no average differences by race or sex. This is a potentially powerful strategy because, if successful, simple comparisons of means can yield strong evidence regarding discrimination.

This is not the place for a review of audit studies, which have many fans and some detractors. For comprehensive reviews and critiques see Heckman (1998) and Fix and Struyk (1993). But there are two important criticisms to note here concerning the limited applicability of audit studies to understanding whether wage gaps reflect discrimination. First, audit studies by nature are limited to an examination of the initial interview and hiring process, and are therefore useful only in assessing discrimination in interviewing and hiring in typically low-paid jobs that do not require reference checks, resume validation, personal recommendations, etc. Second, as has been noted by Heckman (1998), even if audit studies detect some discrimination, this does not imply that the marginal employer discriminates, and it is the marginal employer that, at least in Becker's model of discrimination, determines the wage. Thus, the detection of unequal treatment in audit studies does not necessarily generate a "discriminatory" wage gap.<sup>4,5</sup>

#### *Individual-Level Productivity Data*

---

<sup>4</sup>As Heckman writes, "A well-designed audit study could uncover many individual firms that discriminate, while at the same time the marginal effect of discrimination on the wages of employed workers could be zero" (p. 102).

<sup>5</sup>On the other hand, Black (1995) shows that in a search model, discriminatory tastes on the part of some employers can result in a wage gap, even when the discriminatory employers do not hire any minorities.

A more natural response to the problem of unobserved productivity differences is to study data sets with productivity measures for individuals. As noted before, though, such studies are quite rare because such data are quite rare, and certainly do not exist in a form that permit widely generalizable conclusions to be drawn.

Typifying the idiosyncratic nature of the labor markets for which analysis of individual-level productivity data are available, perhaps the most thorough such analyses have been done for academia. This is a promising occupation to study, of course, because much of the productivity that is thought to be most relevant to pay setting—namely, publications—is readily observed. Furthermore, although not immune from dispute, evidence on citations or the journals in which articles are published can be used to assess not just quantity of output but also quality. Kahn (1995) summarizes a large number of studies focusing on pay discrimination by sex for economics professors, and there are also studies of academics in other fields as well.

A second source of data on individual productivity arises—again, idiosyncratically—when workers are paid both time rates and piece rates. Foster and Rosenzweig (1993) consider what one can learn from such data. On the surface, it would seem that with data on the same worker paid time rates and piece rates for doing the same kind of work, a simple test for taste discrimination would arise, since piece-rate work should capture differences in productivity. Nonetheless, as Foster and Rosenzweig show, one must take great care in attributing wedges between relative wages and relative productivity to any particular form of discrimination. Let  $w$  be the time rate wage,  $F$  be a dummy variable for the sex of the individual, and  $\mu$  be actual productivity measured by piece-rate earnings. Then in the regression

$$w = F\beta + \gamma\mu + \varepsilon ,$$

a negative estimated coefficient on the dummy for female would appear to pick up employer taste discrimination.<sup>6</sup>

---

<sup>6</sup>The vector of observable characteristics of individuals can be generalized to include other covariates in addition to a sex dummy.

Foster and Rosenzweig point out, though, that if there are information problems in labor markets, this may not be a good test. They demonstrate this by an example. Suppose there are four types of workers in a population, men with productivity of either 2 or 4, and women with productivity of either 1 or 3. But suppose that employers cannot distinguish the productive from the unproductive men, or the productive from the unproductive women, but only know the average productivity of men and women. In this case, employers pay expected productivity, so men earn 3 and women earn 2. Now suppose we run a regression of pay on actual productivity and on sex. What will the coefficients be? Sex will perfectly explain the observed pay differences, so the coefficient on actual productivity will be zero, and that on sex will be one, even though there is no taste discrimination. So we cannot tell, from a regression like this, whether we are observing taste discrimination or statistical discrimination.

The “right” way to do the test is to run the regression using expected, rather than actual, productivity, as in

$$w = F\beta + \gamma\mu^* + \varepsilon ,$$

where  $\mu^*$  is expected productivity. Now, in the above example, the coefficients on sex and expected productivity will be zero and one, respectively, indicating (correctly) that there is no taste discrimination.

So if the coefficient on the indicator for females goes to zero in this regression, then we know that we were observing statistical discrimination in the first regression.

We typically do not have information on expected productivity. But if we assume that employers use all available information in forming this expectation, then we know that

$$\mu = \mu^* + u ,$$

where  $u$  is a random error that is uncorrelated with  $\mu^*$ . Thus, we can think of the problem of estimating the first equation as a measurement error problem. The true structural equation is for  $\mu^*$ , but we have an error-ridden proxy  $\mu$ .<sup>7</sup> Thus, we have to instrument for  $\mu$  with variables that are correlated with  $\mu^*$ , but

---

<sup>7</sup>Note that this is a little counter-intuitive, since the realization of true productivity is considered the error-ridden variable; in a model in which employers base pay on expected productivity, however, this is the right way to think about things.

uncorrelated with the measurement error  $u = \mu - \mu^*$ . Otherwise the estimate of  $\gamma$  is biased downward, and if women are on average less productive, the estimated discriminatory sex differential in pay that arises from something other than statistical discrimination is overstated.

As the authors point out, there are three requirements for such instruments. First, they must be correlated with  $\mu^*$ . Second, they must be known to employers; in this case, they are contained in the information set of employers and are therefore orthogonal to  $u$ . Third, they must not themselves appear in the wage equation, implying that they must not be characteristics that are subject to taste discrimination. Satisfying the first two is in fact quite easy, as variables like height, age, education, etc. are available. It turns out that in their data (focusing on the results for the Philippines), if we accept these assumptions, there is no longer any evidence of taste-based discrimination leading to lower wages for women.<sup>8</sup>

The main contribution of this paper in the context of this chapter, however, is to point out that even in the unusual case where we have a measure of actual productivity, care must be exercised in using the simple wage regression approach to test for particular forms of discrimination.<sup>9</sup> If Foster and Rosenzweig's interest in measuring discrimination had been limited to testing for the existence of taste-based wage discrimination, an alternative test using their data would have been available. In particular, one can jointly estimate equations where an observed time-rate wage is regressed on a female dummy in one equation, and the observed piece-rate productivity is regressed on a female dummy in a second equation:

$$w = F\beta + v$$

$$\mu = F\lambda + \eta .$$

---

<sup>8</sup>In an extension, Foster and Rosenzweig (1996) allow for the possibility of comparative advantage that differs across sex and affects the allocation of workers to job tasks. The results are again consistent with statistical discrimination.

<sup>9</sup>A similar problem arises when a data set provides a performance rating. See Neumark (1999) for an application with U.S. data. This is potentially relevant to other studies utilizing performance ratings to try to control for productivity differences, such as Holzer (1990) and Medoff and Abraham (1980).

The error terms in these equations are correlated with the female dummy since  $v = \mu^* + \varepsilon$  and  $\eta = \mu^* + u$ , and  $\mu^*$  is correlated with  $F$ . As a result, the estimated coefficients on the female dummy in both equations will be biased, but they will be biased by exactly the same amount. Therefore, any estimated difference between  $\beta$  and  $\lambda$  can be interpreted as arising from taste-based discrimination rather than statistical discrimination, and failure to find any difference would be inconsistent with taste discrimination.<sup>10</sup> This approach to using wage and productivity information parallels our technique for testing for wage discrimination using matched employer-employee data.

### **III. Using Matched Employer-Employee Data to Test for Wage Discrimination**<sup>11</sup>

#### *The Test*

In order to motivate the approach to testing for wage discrimination that is enabled by matched employer-employee data, we begin with a simple model illustrating the relationship between wages and productivity under perfect competition. Consider an economy consisting of plants that produce output  $Y$  with a technology that utilizes two different types of perfectly substitutable labor inputs,  $L_1$  and  $L_2$ . The production function of these plants is

$$Y = F(L_1 + \varphi L_2) ,$$

where  $\varphi$  is the marginal productivity of  $L_2$  relative to  $L_1$ . These plants are assumed to operate in perfectly competitive spot labor markets, and labor supply is assumed to be completely inelastic. The price of output  $Y$  is normalized to equal one. Wages of workers of types  $L_1$  and  $L_2$  are  $w_1$  and  $w_2$ , respectively. Define the relative wage rate ( $w_2/w_1$ ) to be  $\lambda$ . Given this setup, the proportional mix of the two types of labor in each plant will be determined by the relationship between  $\varphi$  and  $\lambda$ . If  $\varphi = \lambda$ , then under profit maximization or cost minimization plants will be indifferent to the proportional mix of the two types of

---

<sup>10</sup>In the example considered earlier, if we separately regressed wages on sex, and productivity on sex, we would find a coefficient of  $-1$  in each regression, and correctly conclude that there is no taste discrimination.

<sup>11</sup>The framework described here builds on our earlier work in Hellerstein and Neumark (1999) analyzing Israeli manufacturing data, although the data set used in that work is not matched employer-employee data, but is collected from employers only. Cox and Nye (1989) implement a similar approach using employer data from the nineteenth-century French textile industry.

labor in the plant. If there is a wedge between the relative marginal product and relative wage so that  $\varphi \neq \lambda$ , then profit-maximizing or cost-minimizing plants will be at a corner solution, hiring either only workers of type  $L_1$  (if  $\varphi < \lambda$ ) or only workers of type  $L_2$  (if  $\varphi > \lambda$ ). The only equilibrium in this model is when wages adjust so that  $\varphi = \lambda$ , and plants are indifferent between the two types of labor.

Evidence that  $\varphi \neq \lambda$  is inconsistent with the assumption that we are observing profit-maximizing or cost-minimizing plants in a competitive spot labor market.<sup>12</sup> Our approach can be interpreted as providing empirical tests of this characterization of labor markets. We estimate variants of the plant-level production function simultaneously with plant-level wage equations in order to obtain estimates of parameters corresponding to  $\varphi$  and  $\lambda$  for various types of workers. We interpret cases where we cannot reject the equality of  $\varphi$  and  $\lambda$  as evidence consistent with competitive spot labor markets. Cases in which we reject the equality of  $\varphi$  and  $\lambda$  indicate some deviation from this characterization of labor markets, such as would occur in the presence of taste-based discrimination.<sup>13</sup>

#### *Data*

This approach can be implemented using matched data on employers and employees, when the resulting matched data set provides information on demographic characteristics of the employer's workforce, labor costs, and the output and input measures needed to estimate standard production functions. The matched employer-employee data set that we originally used for this purpose is the Worker Establishment Characteristics Database (WECD). The WECD, constructed at the U.S. Census Bureau, links information for a subset of individuals responding to the long form of the 1990 Decennial Census of Population to information about their employers in the 1989 Longitudinal Research Database (LRD), with matches based on the physical location and detailed industry of the worker's reported place of work and similar information on establishments.

---

<sup>12</sup>Labor supply could be less than completely inelastic; as long as market wages remain above reservation wages, the conclusions are unchanged.

<sup>13</sup>This can be couched in the context of Section II. If the production function there is  $f(M + \varphi F)$ , then the first order conditions are  $MP_L = w_M$  and  $\varphi MP_L = w_F + d$ . Dividing through, we get  $MP_F/MP_M = \varphi = (w_F + d)/w_M$ . So if  $d = 0$ ,  $w_F/w_M = \varphi$ , and if  $d > 0$ ,  $w_F/w_M < \varphi$ .



In particular, the Census Bureau maintains a complete list of all manufacturing establishments operating in the United States in a given year (the Standard Statistical Establishment List, or SSEL), along with location and industry information for these establishments that is similar to the data available for workers. Thus, it is possible to assign all plants in the United States to an industry-location cell. The WECD is constructed by first selecting all manufacturing establishments in operation in 1990 that are unique in an industry-location cell. Then all workers who are located in the same industry-location cell as a unique establishment are matched to that establishment.

All that the SSEL provides is an establishment list and some cursory information such as total employment and total annual payroll. To obtain the data on a worker's employer needed to estimate production functions, the employer-employee pairs are matched in a second stage to plant-level data in the LRD. The LRD is a compilation of plant responses to the Annual Survey of Manufacturers (ASM) and Census of Manufacturers (CM). Data in the LRD are of the sort typically used in production function estimation, such as output, capital stock, materials expenditures, and number of workers. In addition, the LRD contains information on total salaries and wages and total non-salary compensation paid by the plant in a given year (McGuckin and Pascoe, 1988). Since worker earnings and labor force information in the Decennial Census of Population refer to 1989, we match the worker data to the 1989 plant data in the LRD. To increase the representativeness of the sample of workers in each plant, we require plants to have at least 20 employees in 1989 (as reported in the LRD), and at least 5% of their workforce contained in the WECD.

The resulting WECD sample contains data on 3,102 plants and 129,606 workers. The average plant has 353 employees, and on average 12% of a plant's workforce is matched to the plant. Further details on the construction of the WECD are given in Hellerstein, et al. (1999) and Troske (1998). In our view, a major advantage of the WECD relative to idiosyncratic data sets with productivity measures is that it provides a relatively large data set that is at least partially representative of a major sector of the U.S. economy.

However, some limitations are worth noting here, in part because more recent efforts at constructing matched employer-employee data sets address some of these limitations. First, the WECD is restricted to the manufacturing sector, since the production function data come from manufacturing surveys. Second, and most important, because possible matches are restricted to establishments that are unique in their location and industry cell, large manufacturing establishments are over-represented; this follows because large establishments are more likely to be unique to industry-location cells.<sup>14</sup> As a counter-example that arises when extending the analysis beyond manufacturing, but continuing to use the industry-location cell to match, small retail establishments, which tend to be located in malls, are unlikely to be unique to an industry-location cell, and hence would be badly under-represented. For similar reasons, nonurban establishments are over-represented. Third, while a large data set results, the WECD nonetheless covers a small fraction of the workforce (about 5% of manufacturing workers who respond to the Census Long-Form), and a small fraction of manufacturing establishments (about 1% of those in the SSEL). Finally, two general limitations of the WECD, and other related matched employer-employee data sets for the United States described below, are (1) only a cross-section is currently available, and (2) only a sample of workers from each establishment is matched to the establishment.

### *Methods*

To estimate parameters corresponding to  $\varphi$ , we estimate a translog production function in which the value of output  $Y$  is a function of capital  $K$ , materials  $M$ , and a quality of labor aggregate  $QL$ . In logs, this is

$$\ln(Y) = \ln(A) + \alpha \ln(K) + \beta \ln(M) + \gamma \ln(QL) + g(K, M, QL) + \varepsilon ,$$

where  $g(K, M, QL)$  is the second-order terms in the production function.

For each plant in our data set, the WECD provides demographic information on a sample of workers. We assume that in the quality of labor aggregate  $QL$ , workers with different demographic characteristics are perfectly substitutable inputs with potentially different marginal products. For

---

<sup>14</sup>There is also an issue of measurement error in reported industry creating incorrect matches (Bayard, 2001).

example, if workers are distinguished only by sex, with the relative marginal productivity of women  $\varphi_F$ , then  $QL$  would be defined as

$$QL = L \cdot \{1 + (\varphi_F - 1) \cdot (F/L)\} ,$$

where  $L$  is the total number of workers in the plant and  $F$  is the number of women in the plant.

Substituting  $QL$  into the general production function, we obtain a production function from which we can estimate  $\varphi_F$ , using plant-level data on output, capital and materials inputs, and the number of workers and sex composition of the workforce.

We have used these data to estimate expanded versions of the type of production functions outlined above, with the difference that rather than distinguishing between only two types of workers, we distinguish between workers based on: sex; race (black and non-black); marital status (ever married); age (divided into three broad categories—under 35, 35-54, 55 and over); education (distinguishing those with at least some college); and occupation (divided into four groups—(1) operators, fabricators, and laborers (unskilled production workers), (2) managers and professionals, (3) technical, sales, administrative, and service, and (4) precision production, craft, and repair). As an example that helps in the ensuing discussion, suppose we use information on sex (male vs. female) and race (black vs. nonblack). Then the expanded expression for  $QL$  is

$$QL = L \cdot \{1 + (\varphi_{WF} - 1) \cdot (WF/L) + (\varphi_{BM} - 1) \cdot (BM/L) + (\varphi_{BF} - 1) \cdot (BF/L)\} ,$$

where  $WF$  is the number of white female workers,  $BM$  the number of black males, and  $BF$  the number of black females in the establishment, with corresponding subscripts on the  $\varphi$ 's.

Once we incorporate multiple characteristics of the workforce, some complicating issues arise. First, a full categorization of workers by the various characteristics listed above (sex, race, marital status, etc.) would require 192 possible combinations of these demographic characteristics. Both the estimation of the share of the workforce in each of these cells, as well as the estimation of the corresponding parameters, would be impractical.

To reduce the dimensionality of the problem, we often use two types of restrictions on the form of  $QL$ . First, we restrict the relative marginal products of two types of workers within one demographic

group to be equal to the relative marginal products of those same two types of workers within another demographic group. For example, the relative productivity of black women to black men is restricted to equal the relative marginal productivity of white women to white men.<sup>15</sup> Similarly, the race difference in marginal productivity is restricted to be the same across the sexes. Second, we restrict the proportion of workers in an establishment defined by a demographic group to be constant across all other groups; for example, we restrict blacks to be equally represented in all occupations, education levels, marital status groups, etc. Upon imposing these restrictions, the previous equation for  $QL$  reduces to

$$QL = L \cdot \{1 + (\varphi_F - 1) \cdot (F/L)\} \cdot \{1 + (\varphi_B - 1) \cdot (B/L)\} ,$$

where  $B$  is the number of black workers, thus eliminating in this case the parameter  $\varphi_{BF}$ . The same restrictions are imposed in estimating the wage equations described below. However, in the empirical work we examine the robustness of the results to relaxing these restrictions along a number of important dimensions identified in previous work (e.g., allowing different marriage differentials for men and women, corresponding to differences in estimated marriage-related wage premia). In such a case there two different implied sex differences, one for married men and women and one for unmarried men and women.

A second issue is that in the production function based on  $QL$  above, labor of each type is perfectly substitutable, as each type enters linearly. It is possible, though, that workers of different types are imperfect substitutes. This is most likely to occur, perhaps, along occupational lines. We have therefore also estimated specifications treating workers in production and non-production occupations as imperfectly substitutable.

As this discussion indicates, many of the limitations of this framework can be addressed by properly generalizing the framework, although there are clear tradeoffs—as we would expect—between restrictions imposed and difficulty of estimation, particularly given sample size constraints. But for the purposes of this chapter, the key is that under many conditions we can estimate parameters (or sets of

---

<sup>15</sup>The analysis is actually done for blacks and nonblacks.

parameters) such as  $\varphi_F$ , the productivity of women relative to men, and  $\varphi_B$ , the productivity of blacks relative to whites.

Because we have information on plant labor costs, we also specify and estimate plant-level earnings equations. These plant-level earnings equations represent the aggregation of individual-level earnings equations over workers employed in a plant, and hence are the plant-level counterparts to the individual-level wage regressions that motivate this research, yielding estimates of parameters  $\lambda_F$  and  $\lambda_B$  (in the example just used), which can be compared to the relative productivity estimates. By simultaneously estimating the production functions and earnings equations at the plant level, we can compare the relative marginal products and relative wages of workers distinguished by various demographic characteristics. While there may be some unobservables in the production function and wage equations, as long as we estimate both of these at the plant level, any biases from these unobservables ought to affect the estimated productivity and wage differentials similarly, at least under the null hypothesis, thus minimizing or eliminating their impact on tests of the equality of relative marginal products and relative wages, and hence isolating the role of discrimination in generating wage differentials.

### *Evidence*

While this framework provides a straightforward and explicit test for wage discrimination, it still leaves open a number of possible variations, including the functional form for the production function, the restrictions imposed on the quality of labor term ( $QL$ ), whether to instrument for potentially endogenous inputs, etc. Regardless, though, it turns out that the estimates are robust.

First, with regard to race, the estimates never provide any statistically significant evidence of wage discrimination. Moreover, the estimates of  $\varphi_B$  and  $\lambda_B$  tend to slightly exceed one. With respect to wages, this conflicts with standard wage equation estimates showing that blacks tend to earn less than similar whites. This discrepancy probably arises for two reasons, which are instructive as to some potential pitfalls with matched employer-employee data. First, when we estimate a wage equation for the WECD including fixed plant effects, the estimated wage differential between blacks and whites in the

WECD is a bit larger ( $-.08$ ) than in the specification without fixed effects ( $-.05$ ), indicating that within plants blacks earn less than whites, but that blacks work in slightly higher-paying plants.<sup>16</sup> Thus, because the test for discrimination we have developed using the WECD relies on variation across plants, the earnings of blacks relative to whites are biased upward. While, as just described, the wage differential associated with race can be estimated within plants, the productivity differential can only be estimated across plants. Second, comparing sample means for the matched sample and manufacturing workers in the Census overall reveals that blacks are under-represented in the WECD. If this arises as undercounting of blacks relative to whites within plants, then this will tend to bias estimates of  $\varphi_B$  and  $\lambda_B$  away from one, as any productivity or wage differential associated with having more black workers is attributed to a smaller share of blacks than are probably in the establishment.

Thus, the unusual estimates for blacks may stem from the reliance on across-plant variation for identification, and on non-representativeness of the employees who are matched. There is no obvious solution to the first problem within this framework (although, as discussed later, panel data are potentially helpful), while efforts to construct better, more representative matched employer-employee data sets are described below.

Turning to the results regarding sex discrimination, we begin by noting that the “problem” to be explained or explored is much more severe with respect to sex than with respect to race. If we simply use the worker data to estimate standard log wage regressions, the estimated race gap is 5 log points, while the estimated sex gap is 35 log points. The production function and wage equation results, many of which are collected in Table 1, tell a very consistent story. The baseline estimate of  $\varphi_F$  is 0.84, which is significantly less than one, while the estimate of  $\lambda_F$  is 0.55, also significantly less than one, but much smaller than  $\varphi_F$ . Moreover, the p-value for the test of equality of these two parameters is effectively zero, indicating that the null hypothesis of no discrimination is rejected, in favor of the hypothesis of wage

---

<sup>16</sup>See also Carrington and Troske (1998) for similar evidence.

discrimination against women.<sup>17</sup> Across alternative specifications the estimate of  $\lambda_F$  is very robust. The estimate of  $\varphi_F$  varies somewhat more, between a range of about .71 to 1.01, but the test of the null of no discrimination is rejected in all cases except for managerial/professional occupations when we relax the dimensionality restrictions by sex and occupation. The estimates suggest that while in manufacturing in the United States women may be less productive than men, the pay gap is considerably larger, consistent with wage discrimination.

### *Limitations*

There are some inherent limitations to the analysis we have been able to conduct using the WECD. Some of these have been discussed already, including non-representativeness, a relatively small fraction of matched employers and employees, and the cross-sectional nature of the data set and therefore the identification strategy. An issue related to the relatively small fraction of matched workers is measurement error. In particular, our specifications rely on knowing the percentages of workers with various demographic characteristics in each plant, yet these are actually estimated from the sample of workers that are matched to the plant. We have used simulation methods (see Hellerstein, et al., 1999, Section VIII) to assess the biases caused by this measurement error, generally concluding that the effect of measurement error is to bias our estimates toward finding no discernable productivity or wage differentials across workers, and toward finding no differences between the relative productivity and wage estimates for a given type of worker (since estimates of parameters that are further from one have larger absolute biases toward one). Thus, the power of the tests of the equality of wage and productivity differentials is somewhat reduced because of measurement error, strengthening our rejection of this

---

<sup>17</sup>Of course, how this result is interpreted depends in part on the source of variation in the share of the plant's workforce that is female. Section II showed how this variation can be driven by differences in discriminatory tastes, in which case all employers adjust their share female so that their disutility from female workers just matches the wage differential relative to the productivity differential that we estimate. Another possibility we have explored (Hellerstein, et al., 1999) is that the share female is correlated with other variables, such as technological innovation. A separate survey of manufacturing technology for some plants that can be identified in our data (in high-technology industries) suggests that, if anything, the share female is lower in plants using advanced technologies. This might lead to downward bias in the estimated extent of discrimination, as the apparent higher productivity of men is in part attributable to more advanced technology.

equality. Nonetheless, these simulations do not give a precise answer, and a data set with a much higher fraction of matched workers would be useful.

#### **IV. Other Uses of Matched Employer-Employee Data to Learn About Labor Market**

##### **Discrimination**

###### *Exploring the Becker Model of Discrimination*

The Becker model of employer discrimination is often interpreted as ruling out wage discrimination of the type apparently detected by the residual wage gap method, and (more convincingly, we believe) by the production function approach just outlined. Becker shows that under sufficiently competitive conditions discriminatory employers will fail to thrive, and will eventually be competed out of existence. Some have leapt from this theoretical result to the claim that, since the sex-wage differential has persisted, it must reflect unobserved productivity differences rather than discrimination (Fuchs, 1988; O’Neill, 1994).

The argument that competition undermines discrimination is based on two implications of the model. First, there is a static implication that discrimination is likely to exist only where there is product market power, and conversely that product market competition hinders discrimination. Second, there is a dynamic implication that even if there are firms with product market power that can “afford” to discriminate in the short run, over the long run market forces eliminate the discrimination. However, whether or not market forces are expected to undermine discrimination depends on the nature of product market competition, barriers to entry, transferability of assets, and the form of employers’ discriminatory tastes. In particular, if employers’ tastes are nepotistic rather than discriminatory, then discrimination will not be eliminated by competition in the market for firms, although product market competition may still suffice.<sup>18</sup> In our view, the conditions for competition to eliminate discrimination have been glossed over by those who interpret the Becker model as implying that discrimination cannot persist. Rather, we think it essential to try to test these implications of Becker’s model empirically.

---

<sup>18</sup>See Becker (1971, Chapter 3) for a detailed discussion of the conditions under which competition should erode discrimination.



Individual-level wage regression estimates can tell us virtually nothing about the relationship between market forces and discrimination. Matched employer-employee data, in contrast, can be quite informative. In Hellerstein, et al. (2002) we use the WECD to test both the static and dynamic implications of the Becker model, in both cases exploiting the links to employer-level data that arise from the matching.

We begin with a simple test for sex discrimination. If there is no discrimination against women, then there should be no cross-sectional relationship between profitability and the sex composition of the workforce. Any sex difference in wages must reflect only observed or unobserved productivity differentials between men and women, and firms or plants that employ more women should earn no higher profits. Evidence that plants or firms that employ relatively more women earn higher profits, in contrast, would be consistent with sex discrimination. This test, not surprisingly, parallels the production function-based test described in the previous section.

Then, moving on to the implications of the Becker model for the role of market forces, we first test whether—as the Becker model predicts—there is a positive short-run relationship between profitability and the sex composition of the workforce only among plants with product market power, because only among such plants are there positive economic profits that may be exploited to indulge the discriminatory tastes of some employers. Finally, we use longitudinal data on the plants in our data set to test the dynamic implications of Becker’s model, asking whether non-discriminatory plants—those which, according to the cross-sectional evidence, employ more women and earn higher profits—grow faster or are less likely to undergo a change of ownership, compared with discriminatory plants. The combined static and dynamic evidence provides us with a better understanding of sex discrimination in labor markets, and the role of competitive market forces in reducing or eliminating this discrimination, than can be obtained from conventional wage equation approaches to discrimination, or even from more convincing cross-sectional tests for discrimination (such as audit studies or our production function-based tests), which cannot say anything about the dynamic implications of Becker’s model.

The first test relies solely on the WECD, described above. The only new feature is the construction of a measure of performance to use as a proxy for profitability. One measure we use is operating income (sales + inventory accumulation – labor costs – material costs) divided by inventory-adjusted sales (that is, current year sales plus the change in inventories), while a second measure also subtracts out from the numerator purchased business services or overhead costs (such as purchased electricity, legal services, advertising, and repairs). These performance measures used as proxies for profitability are essentially price cost-margins (see, for example, Domowitz, et al., 1986).

In examining the relationship between profitability and the proportion female, we include as controls other demographic characteristics of the workforce, and plant or firm characteristics that are likely to be associated with profitability, such as age, industry, and measures of market power. Estimates of the cross-sectional relationship between profitability and the proportion female indicate that irrespective of variations in the sample, variable construction, and the use of plants or firms, the proportion female in the workforce is significantly positively related to profitability.

We next consider the implication of the Becker model that product market power can inhibit the influence of competitive forces in eliminating discrimination (although competition in the market for firms may still exert such influence over the long run). This implies that for plants with considerable product market power there will be a positive relationship between employment of women and profitability, whereas for plants with little market power, the relationship will be non-existent or at least weaker. The results indicate that for plants with a smaller market share there is a small and statistically insignificant relationship between profitability and the proportion female. But this relationship becomes positive and statistically significant at higher output shares.<sup>19</sup> This positive relationship between the proportion female and profitability among plants with high output share is consistent with the model.<sup>20</sup> In

---

<sup>19</sup>Kawaguchi (2003) reports related and similar evidence based on data from a census of Japanese firms meeting minimum thresholds for employment and capital.

<sup>20</sup>These results are also consistent with tabulations from the 1940 Census reported by Becker, indicating that relative to firms operating in competitive industries, firms operating in monopolistic

plants that operate in a competitive output market, there is no performance advantage to hiring women; since discriminatory employers are unable to indulge their tastes for discrimination, wages in this sector are quickly bid to equality with marginal products. But in (typically larger) plants where there is market power and where plants can discriminate if they wish, at least in the short run, those that do not discriminate and hire more women—who cost less—achieve better performance.

We interpret our cross-sectional results as suggesting that among the set of plants with product market power—which are typically larger plants—some exploit the supra-competitive profits stemming from their product market power to indulge in discriminatory behavior against women. Given that wages are set by the market, we do not expect the wage gap between men and women to differ between the less- and more-competitive sectors. Rather, a larger gap between relative wages and relative marginal products in the less-competitive sector would have to reflect higher relative productivity of women in the less-competitive sector, possibly because of technological differences (the perfect substitutes case), or because fewer women are employed than would be the case in the absence of discrimination (the imperfect substitutes case) in this sector. The data are consistent with these predictions. In our earlier research, we found no productivity gap between men and women among larger firms, but that in smaller firms women were less productive than men. Moreover, wage regressions using plant-level data indicate that the relative wage of women is not particularly low among plants in the less-competitive sector that employ a high proportion female.

The cross-sectional results suggest that plants or firms can indulge their discriminatory tastes only when they enjoy product market power. Nonetheless, market forces may reduce or eliminate discrimination as the more profitable plants or firms among those with market power grow relative to the less profitable ones, or as discriminating employers are bought out by non-discriminating employers. We look at evidence regarding this dynamic prediction of the Becker model for plants with relatively high output share (located in the top quartile of the distribution of this market power measure). For these

---

industries appear to engage in discriminatory behavior (1971, Table 2, p. 48); Becker's results, however, refer to hiring of non-whites.

plants, we estimate regressions for growth of both employment and shipments, and probits for ownership change, as functions of, among other things, the proportion female in a plant's workforce.

When we look at growth, the estimated coefficients on the proportion female are small and statistically insignificant. There is therefore no evidence that sex discrimination in wages is reduced or eliminated by market forces that cause non-discriminatory plants to expand relative to discriminatory plants. The estimated effect of higher relative employment of women in the base year on the probability of changing ownership between 1990 and 1995 is negative, consistent with plants employing fewer women being more likely to change ownership, as predicted by Becker's model. However, the relationship is not statistically significant. Overall, then, these results provide little evidence of market forces that bid away the sex discrimination reflected in the cross-sectional relationship between profitability and the proportion female for plants with product market power. We emphasize, however, that these results on growth and ownership change are derived from five-year changes, and it is possible that this is not a long enough period of time to be able to measure the effect of competitive market forces on discriminatory practices. On the basis of this evidence, though, there is not an empirical case for concluding that product market competition alone will suffice to eliminate wage discrimination.

Levine and Leonard (2002) also use matched employer-employee data to examine theories of discrimination. Specifically, they have assembled a large longitudinal data set matching up employees to establishments of a large retail firm, coupled with Census data on the demographics of the population near each establishment. They use these data to examine theories of co-worker and customer discrimination, as well as to try to estimate the effects of workforce diversity on performance (sales).

#### *Sex Segregation and the Sources of the Sex Gap in Wages*

Matched employer-employee data are also critical in providing a full description of segregation in the labor market, and in estimating the contribution of various dimensions of segregation to wage gaps between demographic groups. A prominent example of literature on this topic concerns the wage gap between men and women in U.S. labor markets. Women have consistently earned lower wages than men in U.S. labor markets, although this gap has narrowed in recent decades (Blau, 1998). Previous research

has largely focused on the impact on the wage gap of sex segregation by occupation (e.g., Macpherson and Hirsch, 1995) and by industry (e.g., Fields and Wolff, 1995). However, a smaller literature has focused on the independent contribution of the segregation of men and women into different employers (e.g., Blau, 1977), and segregation among occupations or jobs within establishments, or “job-cell” segregation (Groshen, 1991).

While the latter two dimensions of segregation appear quite important empirically, it is difficult to find systematic evidence on establishment or job-cell segregation. Indeed, there is no empirical work on this issue in the earlier literature using large data sets representative of a wide array of industries, because of the absence of such data sets containing detailed demographic information for multiple workers in the same establishment. Instead, Blau’s and Groshen’s studies of the effects of establishment and occupation-establishment segregation examined unusual, quite narrow data sets, with Groshen using surveys of wages for a subset of occupations in five specific industries included as part of the Bureau of Labor Statistics (BLS) Industry Wage Surveys, and Blau using the BLS Area Wage Surveys. The focus in these studies on a handful of industries or occupations provides something closer to a set of case studies, with the lack of representativeness limiting their usefulness in assessing the forces at work in generating the sex wage gap in the United States.

In contrast, the methods we have developed to match employer and employee data using the Census data sources can provide much broader and more nationally representative data sets with which to estimate the contributions of sex segregation by industry, occupation, and occupation-establishment cell (job cell) to the sex wage gap. Whereas the work described earlier required the estimation of production functions, and hence was restricted to manufacturing establishments, to study segregation we have constructed an extended version of the WECD. Like the WECD, this data set uses the U.S. Census Bureau’s SSEL to identify the employers of individuals who responded to the long form of the 1990 Decennial Census. However, this new data set (the New Worker-Establishment Characteristics Database, or NWECD) includes workers and establishments from all sectors of the economy and all regions; furthermore, because the production function “ingredients” are not needed, there is no requirement to

match to the LRD. As described in Bayard, et al. (2003), this results in the NWECD containing information on over 600,000 workers in nearly 33,000 establishments, once restrictions similar to those imposed in the construction of the WECD, and other restrictions needed to make the analysis comparable to the existing literature, are imposed. Nonetheless, the constraints imposed by matching employees to employers based on unique industry-location cells remain, leading to over-representation of manufacturing establishments and nonurban establishments. Additional work that attempts to address these limitations is described below.

We use the NWECD to provide new estimates of the roles of various dimensions of sex segregation in generating sex differences in wages. Although in some respects our evidence may be viewed as complementary to that in the earlier studies, in our view the NWECD, while having some shortcomings, is clearly better suited to characterizing the effects of sex segregation in U.S. labor markets.<sup>21</sup>

The empirical work proceeds by estimating wage regressions for men and women including a dummy variable for women, and the percentages female in the individual's occupation, industry, establishment, and job cell (or establishment-occupation cell). These regressions yield two sets of results. The first is a decomposition of the share of the sex gap in wages attributable to each dimension of segregation, and the second is the unexplained sex gap that remains after controlling for this segregation, or the "within-job cell" sex gap in wages. Our results indicate that a sizable fraction of the sex gap in wages is accounted for by the segregation of women into lower-paying occupations, industries, establishments, and occupations within establishments. For example, using the most-detailed occupational classification, 4% of the gap is attributable to occupational segregation, 11% to industry segregation, 18% to establishment segregation, and 24% to segregation across occupations within establishments. However, a substantial part of the sex wage gap remains attributable to the individual's sex, in this example 40%, with other estimates ranging up to around 50%.

---

<sup>21</sup>Datta Gupta and Rothstein (2001) replicate this study using matched employer-employee data from Denmark.

These magnitudes are significant because they inform policy. In particular, if within the narrowly-defined occupations that we study the jobs performed by men and women require substantially equal skill, effort, responsibility, and working conditions, yet wages differ by sex, then enforcement of the Equal Pay Act can play a fundamental role in closing the wage gap between men and women. In contrast, if segregation along various dimensions accounts for most of the sex wage gap, then policies along the lines of comparable worth, equal opportunities in employment and promotion, and affirmative action would be central to any further closing of this gap, and stronger equal pay provisions would not be effective. Our findings suggest that stronger enforcement of equal pay legislation could further reduce the wage gap between men and women, perhaps substantially.<sup>22</sup>

Finally, in a related paper (Bayard, et al., 1999) using the same type of wage decomposition analysis, we use the NWECD to determine whether segregation can help explain why racial and ethnic wage gaps are larger for men than for women. In the NWECD, the black-white wage gap for men is  $-0.23$  and the Hispanic-white wage gap is  $-0.24$ ; for women, the black-white wage gap is  $-0.16$  and the Hispanic-white wage gap is  $-0.13$ . The results indicate that greater segregation between Hispanic men and white men than between Hispanic women and white women accounts for essentially all of the higher Hispanic-white wage gap for men. But the estimates indicate that greater segregation between black and white men than between black and white women accounts for only one-third to one-half of the higher black-white wage gap for men. Another way to interpret these results is to conclude that segregation is an important contributor to the lower wages paid to black and Hispanic men compared to white men with similar individual characteristics, although more so for Hispanic men. And while equal pay laws may offer some scope for reducing the black-white wage differential for men, they offer less scope for reducing the Hispanic-white wage differential for men.

## **V. Other Sources of Matched Employer-Employee Data**

### *New Matched Data Sets for the United States*

---

<sup>22</sup>See Bayard, et al. (2003) for an analysis of why these findings differ from Groshen's, who concluded that the within-job-cell sex wage gap was very small.

We hope that the material presented to this point establishes the value of using matched employer-employee data to study labor market discrimination. But we also noted a few important limitations of the WECD (and the NWECD), including the small fraction of establishments matched and non-representativeness. We have recently made progress on developing matched employer-employee data for the United States that rectify these deficiencies.

In particular, we have developed an alternative method to match workers to employers that does not require establishments and workers to be located in unique industry-location cells. Instead, this method relies on matching the actual employer name and address information provided by respondents to the Decennial Census to name and address information available for employers in the SSEL. We refer to this data set as the Decennial Employer-Employee Dataset, or the DEED. This produces a matched data set that is much larger and more representative than the WECD or the NWECD.<sup>23</sup>

When the NWECD was created, the specific name and address files for Long-Form respondents were unknown and unavailable to researchers. Subsequently, we were able to help track down the name and address files and to participate in their conversion from an internal Census Bureau input/output language to a readable format. Because this name and address file had been used solely for internal processing purposes, it did not have an official name, but was informally known as the “Write-In” file.

The Write-In file contains the information written on the questionnaires by Long-Form respondents, but not actually captured in the SEDF. For example, on the Long Form workers are asked to supply the name and address of their employers, while in the SEDF the only information retained is a set of geographic codes (state, county, place, tract, block). The Write-In file, however, contains the geographic codes as well as the employer’s actual business name and address. Because name and address information is also available for virtually all employers in the SSEL, nearly all of the establishments in the SSEL that are classified as “active” by the Census Bureau are available for matching, allowing us to use employer names and addresses for each worker in the Write-In file to match workers to employers in the SSEL. Additionally, because both the Write-

---

<sup>23</sup>Because the WECD contains only manufacturing establishments, while the DEED and the NWECD cover all industries, in the remaining discussion we focus only on comparing the latter two data sets.



In file and the SEDF contain identical sets of unique individual identifiers, we can use these identifiers to link the Write-In file to the SEDF. This procedure potentially yields a much larger matched data set, and one whose representativeness (as well as size) is not compromised by the need to focus on establishments unique to industry-location cells.

The key complication in implementing this procedure arises because names and addresses are not reported identically (or completely) by workers and their employers. Consequently, it is necessary to utilize “fuzzy” matching software that constructs matches out of non-identical records. This is a very complex and detailed process, and is described in detail in Hellerstein and Neumark (2003). Suffice it to say that we developed matching algorithms that err very strongly on the side of caution—avoiding matches with a non-negligible probability of being invalid at the risk of eliminating many valid matches. Along the way, we also hand-checked many thousands of matches to detect problems with the matching algorithm and to fine-tune the algorithm to avoid these problems.

The resulting data set is very large. Out of all 12,143,183 workers in the SEDF who met the basic criteria for sample inclusion, 3,291,213 (approximately 27%) are in the DEED, a substantial improvement over the NWECD, which contains only 7% of possible matches. Turning to establishments, there are 5,237,592 establishments in the SSEL; of these, 972,436 (19%) also appear in the DEED, vs. 137,735 (slightly more than 2.5%) in the NWECD. In addition, the DEED is far more representative of both workers and establishments. The non-representativeness of the NWECD, stemming from the matching algorithm, was most pronounced with respect to industry, with over-representation of workers in manufacturing and under-representation of retail workers. This problem is greatly ameliorated in the DEED. Approximately 25% of all workers in the SEDF are employed in the manufacturing sector, and although this number is somewhat greater in the DEED (33%), it is substantially lower than in the NWECD (49%). Retail workers comprise 20% of all workers in the SEDF, and 17% in the DEED, but only 9% of all NWECD workers. Paralleling this, the distributions of establishments across industries in the DEED and NWECD relative to the SSEL are similar to those in the worker sample in the sense that the DEED is much closer to the SSEL. For example, although there are roughly the same shares of services establishments in all three data sets (28% in the SSEL, 26% in the DEED, and 27% in the NWECD), there is a far greater representation of manufacturing establishments in

the NWECD (29%) than in the SSEL (6%) or the DEED (13%). Examining the distributions of establishments across geographic areas also reveals that the DEED is more representative of the SSEL than is the NWECD. In both the SSEL and the DEED just over 81% of establishments are in an MSA, while this is true for only 61% of NWECD establishments.

Thus, the DEED offers substantial improvements in providing a matched employer-employee data set for the United States that is larger and more representative of the actual population of workers and establishments. We are confident that the DEED will prove very useful in revisiting some of the questions regarding labor market discrimination that we have studied in the past with the WECD and NWECD. In addition, the massive size of the DEED makes it uniquely suited to studying additional questions regarding labor market discrimination, in particular those that require a very fine disaggregation of workers and/or employers. For example, our first paper using the DEED studies the role of labor market segregation based on Hispanic ethnicity and language proficiency (Hellerstein and Neumark, 2003). We have also used the subset of manufacturing establishments in the DEED to replicate the earlier results in Table 1, based on the WECD, regarding sex differences in wages and productivity in this sector (Hellerstein and Neumark, 2004). We have found that the results are very similar, although of course the estimates are considerably more precise.

Another national matched employer-employee data set currently under construction at the U.S. Census Bureau is the Longitudinal Employer-Household Dynamics (LEHD) data. The LEHD is very rich in that it contains observations on all workers in covered establishments (not limited to the 1-in-6 sample of Long-Form respondents) and is longitudinal in nature. However, the LEHD only covers some states (although many of the largest ones). In addition, it is based on firm-level rather than establishment-level data, so that workers can only be accurately matched to establishments when the establishment is not part of a multi-unit firm. However, the LEHD should ultimately permit some of the same analyses described

in this chapter, with additional statistical analyses and testing of hypotheses regarding discrimination enabled by the longitudinal nature of the data.<sup>24</sup>

### *International Matched Data Sets*

To this point, the discussion has focused on research using matched data for the United States. As reviewed in Abowd and Kramarz (1999), employer-employee data sets of some variety or other have been assembled for numerous countries. In contrast to the data for the United States, where the main employer-employee data sets have resulted from matches across different surveys, many of the data sets in other countries are based on surveys that set out to cover both establishments and employers, although others are based on matches. For the most part, these data sets have not been used to address questions regarding labor market discrimination. This is probably in part because such questions are simply far more central to labor economics research in the United States. However, some recent work has implemented our production-based tests for wage discrimination for other countries.

Haegeland and Klette (1999) study Norwegian manufacturing plants using data from Statistics Norway including an annual census of establishments, and “register data files” covering essentially all workers. While there is some non-correspondence between the employer identification numbers in the two data sources, the authors assert that they match workers to employers for over 70% of the manufacturing establishments they study. The empirical methods they use follow those in Hellerstein, et al. (1999) closely. But their results are quite different, pointing to wage and productivity shortfalls of women relative to men in the range of about 18-19%, and no significant evidence against the equality of relative wages and marginal products of women and men.

Similarly, Crépon, et al. (2002) study matched data covering the period 1994-1997 for France, based on the Bénéfice Réels Normaux, an employer-level file, and the Déclarations Annuelles des Données Sociales, an exhaustive employee-level file. Again, the available employer identification code

---

<sup>24</sup>Yet another source of employer-employee data in prior research is Unemployment Insurance records maintained by states (see Abowd and Kramarz, 1999). But these typically provide very limited information on workers and employers, and hence are unlikely, by themselves, to prove very instructive in studying discrimination.

does not permit a perfect match, as 9% of establishments do not appear in the employee data file. One downside to this data set is that the worker information is sparse, for example lacking measures of education level and marital status. Crépon, et al., also use similar methods to Hellerstein, et al. (1999), with slight modifications. Paralleling the results for Norway, these authors also find little substantive evidence of discrimination against women. Their estimates imply that women are less productive (by 11% in manufacturing, and 7% in non-manufacturing), and that the pay differentials, while significantly larger than this (14% and 9%, respectively), are substantively quite close to the productivity differentials.

Finally, as noted earlier we have also implemented these methods using data on Israeli manufacturing plants (Hellerstein and Neumark, 1999). There, too, we find that there is both a pay gap and a productivity gap favoring men over women, but that the gaps are of roughly equal magnitude, failing to provide evidence of wage discrimination.

Thus, to this point, results from implementing this test based on matched employer-employee data for the United States yield strong evidence of discrimination against women. In contrast, for the handful of other countries for which this method has been used, women appear lower paid *and* less productive, and there is generally little evidence of discrimination. There are a number of possible reasons why the results may vary across countries. First and most obviously, given differences in employment laws and institutions, and given cultural differences about the roles of men and women in the home and at work, one might not expect the answer to be the same across countries. Second, and more subtly, even two economies with the same laws, institutions, and cultural norms may yield different equilibria with respect to wage discrimination. Recall that in the Becker model of taste discrimination, it is the marginal firm that determines the equilibrium wage differential between workers. Differences across countries in the composition of firms (most importantly, perhaps, in the distribution of employers' discriminatory tastes), or differences in the relative supplies of female labor across countries, can yield large differences in the characteristics of the marginal employer, and thus large differences in equilibrium wages across men and women.

For example, in more disaggregated analyses using the data for the United States, we find evidence that the relative wage for women is equal to the relative productivity for women who work in small establishments, but in large establishments women are relatively more productive than in small establishments, but are underpaid. Thus, it seems that the market wage differential between men and women is set by small establishments, and large establishments are able to pay that market wage and indulge in a taste for discrimination, presumably because they have market power.

Our full-sample results and those reported for other countries report average estimates across all establishments of relative productivities and relative wages, and it is therefore impossible to tell from these results whether cross-country differences may be due to differences in the identity of the marginal employer, whose tastes determine market wages. Nonetheless, nothing in this argument rules out interpreting the results for the United States as indicating wage discrimination. Rather, the open question is whether the absence of such evidence in a few other countries reflects less discriminatory tastes, differences in the marginal employer, or variation in the legal/institutional setting. It remains to future research to probe the cross-country differences more fully, focusing both on questions regarding data and empirical methods, as well as the economic, legal, institutional, or other factors that could underlie these differences.

Two additional points regarding the international data sets are worth noting. First, in contrast to data for the United States, some of the international data sets (including those for Norway and France discussed above) have the advantage of relying on data covering all workers at an establishment, rather than a sample. Second, these two international data sets, as well as others, include repeated observations on employers and their workforces. This feature can in principle be exploited to control for underlying establishment-level heterogeneity, by asking, for example, whether the average wage paid in an establishment actually falls when the proportion female in the establishment's workforce rises, or whether instead low-wage establishments tend to employ a greater share of women. With cross-sectional data, of course, we cannot distinguish between these two hypotheses.

In the United States, as the matched data are based on the Decennial Census of Population, the best that one can do with respect to constructing panel data is to use observations ten years apart. While this is something we plan to explore using the 2000 Census, the assumption that establishment-level unobservables are constant over such a long time horizon is tenuous. The data used by Haegeland and Klette (1999) consist of a panel of yearly observations over the 1986-1993 period. Yet they choose to use time averages of the data (the “between” estimator) to reduce measurement error, rather than to exploit the longitudinal structure of the data set more fully. The French data used by Crépon, et al. (2002) also have a panel structure, but they, too, rely almost exclusively on the between variation, and only report limited results for the wage equation using differenced estimators. Finally, in other work using the Israeli data, we have attempted to implement versions of our test for wage discrimination using panel data from two years (Hellerstein and Neumark, 1998). We find that changes in the proportion female are not associated with declines in wages, but are associated with declines in productivity (albeit with imprecise estimates), which we interpret as evidence of lower productivity of female employees, coupled with clustering of women in low-wage plants but some constraints on paying women and men differentially within establishments. However, the extension of these methods to panel data is best characterized as remaining largely unexplored to date.

## **VI. Summary and Conclusions**

The recent construction of matched employer-employee data sets has allowed researchers to test in new ways for the presence of market-wide discrimination. Data at the establishment-level are used to estimate marginal productivity differentials between workers, which are in turn compared directly to market wage differentials. This is a formal test for the existence of taste-based employer discrimination that, if sufficiently pervasive, generates a wedge between the relative market wages of groups of workers and their relative productivities. Because it directly tests the theoretical prediction of the taste-based model, this kind of analysis is far more convincing than approaches based on residual wage comparisons, and offers advantages over other methods of testing for discrimination as well.

To this point, results from implementing this test based on matched employer-employee data for the United States yields strong evidence of discrimination against women. In contrast, for the handful of other countries for which this method has been used, women appear lower paid *and* less productive, and there is generally little evidence of discrimination. It remains to future research to probe these differences more fully, focusing both on questions regarding data and empirical methods, as well as the legal, institutional, or other economic factors that could underlie these differences.

In addition to enabling this direct test for discrimination, matched employer-employee data sets also hold the potential to provide researchers with more information on the nature of discrimination and labor market differences between groups more generally, and by extension, therefore, to inform debate over possible policy responses. We fully expect that as the depth and scope of matched employer-employee data sets both in the United States and abroad are expanded, much more will learned from these data about the existence, extent, and nature of labor market discrimination.

## References

- Abowd, John M., and Francis Kramarz. 1999. "The Analysis of Labor Markets using Matched Employer-Employee Data." In Orley Ashenfelter and David Card, eds., Handbook of Labor Economics, Vol. 3B (Amsterdam: Elsevier), pp. 2629-2710.
- Bayard, Kimberly. 2001. "Measurement Error and Inter-Industry Wage Differentials." Unpublished paper, Board of Governors of the Federal Reserve System.
- Bayard, Kimberly, Judith Hellerstein, David Neumark, and Kenneth Troske. 1999. "Why Are Racial and Ethnic Wage Gaps Larger for Men than for Women? Exploring the Role of Segregation Using the New Worker-Establishment Characteristics Database." In John Haltiwanger, Julia Lane, James Spletzer, Jules Theeuwes, and Kenneth Troske, eds., The Creation and Analysis of Employer-Employee Matched Data (Amsterdam: Elsevier Science B.V.), pp. 175-203.
- Bayard, Kimberly, Judith Hellerstein, David Neumark, and Kenneth Troske. 2003. "New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data." Journal of Labor Economics, Vol. 21, No. 4, October, pp. 887-922.
- Becker, Gary S. 1971. The Economics of Discrimination, 2<sup>nd</sup> edition (Chicago: The University of Chicago Press).
- Becker, Gary S. 1985. "Human Capital, Effort, and the Sexual Division of Labor." Journal of Labor Economics, Vol. 3, No. 1 (Pt. 2), January, pp. S33-58.
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." American Economic Review, Vol. 94, No. 4, September, pp. 991-1014.
- Black, Dan A. 1995. "Discrimination in an Equilibrium Search Model." Journal of Labor Economics, Vol. 13, No. 2, April, pp. 309-34.
- Blau, Francine D. 1977. Equal Pay in the Office (Lexington, MA: D.C. Heath and Company).
- Blau, Francine D. 1998. "Trends in the Well-Being of American Women, 1970-1995." Journal of Economic Literature, Vol. 36, No. 1, March, pp. 112-65.
- Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." Journal of Human Resources, Vol. 8, No. 4, pp. 436-55.
- Carrington, William and Troske, Kenneth. 1998. "Interfirm Racial Segregation." Journal of Labor Economics, Vol. 16, No. 2, April, pp. 231-60.
- Cotton, Jeremiah. 1988. "On the Decomposition of Wage Differentials." Review of Economics and Statistics, Vol. 70, pp. 236-43.
- Cox, Donald, and Nye, John V. 1989. "Male-Female Wage Discrimination in Nineteenth-Century France." Journal of Economic History, Vol. XLIX, December, pp. 903-20.
- Crépon, Bruno, Nicolas Deniau, and Sébastien Pérez-Duarte. 2002. "Wages, Productivity, and Worker Characteristics: A French Perspective." Unpublished Paper, Centre de Recherche en Économie et Statistique and Institut National de la Statistique et des Études Économiques.



Darity, William A., and Patrick L. Mason. 1998. "Evidence on Discrimination in Employment: Codes of Color, Codes of Gender." Journal of Economic Perspectives, Vol. 12, No. 2, Spring, pp. 63-90.

Datta Gupta, Nabanita, and Donna S. Rothstein. 2001 "The Impact of Worker and Establishment-Level Characteristics on Male-Female Wage Differentials: Evidence from Danish Matched Employee-Employer Data." BLS Working Paper 347.

Domowitz, Ian, R. Glenn Hubbard, and Bruce C. Peterson. 1986. "Business Cycles and the Relationship Between Concentration and Price-Cost Margins." Rand Journal of Economics, Vol. 17, No. 1, pp. 1-17.

Ferber, Marianne A., and Carole A. Green. 1982. "Traditional or Reverse Sex Discrimination? A Case Study of a Large Public University." Industrial and Labor Relations Review, Vol. 35, No. 4, pp. 550-64.

Fields, Judith, and Edward N. Wolff. 1995. "Interindustry Wage Differentials and the Gender Wage Gap." Industrial and Labor Relations Review, Vol. 49, No. 1, October, pp. 105-20.

Fix, Michael, and Raymond J. Struyk, eds. 1993. Clear and Convincing Evidence: Measurement of Discrimination in America (Washington, DC: The Urban Institute).

Foster, Andrew D., and Mark R. Rosenzweig. 1993. "Information, Learning, and Wage Rates in Low-Income Rural Areas." Journal of Human Resources, Vol. 28, No. 4, Fall, pp. 759-90.

Foster, Andrew D., and Mark R. Rosenzweig. 1996. "Comparative Advantage, Information and the Allocation of Workers to Tasks: Evidence from an Agricultural Labour Market." Review of Economic Studies, Vol. 63, No. 3, July, pp. 347-74.

Fuchs, Victor R. 1988. Women's Quest for Economic Equality (Cambridge, MA: Harvard University Press).

Goldin, Claudia, and Cecilia Rouse. 2000. "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians." American Economic Review, Vol. 90, No. 4, September, pp. 715-41.

Haegeland, Torbjørn, and Tor Jakob Klette. 1999. "Do Higher Wages Reflect Higher Productivity? Education, Gender and Experience Premia in a Matched Plant-Worker Data Set." In John Haltiwanger, Julia Lane, James Spletzer, Jules Theeuwes, and Kenneth Troske, eds., The Creation and Analysis of Employer-Employee Matched Data (Amsterdam: Elsevier Science B.V.), pp. 231-60.

Heckman, James J. 1998. "Detecting Discrimination." Journal of Economic Perspectives, Vol. 12, No. 2, Spring, pp. 101-16.

Hellerstein, Judith K., and David Neumark. 1998. "Wage Discrimination, Segregation, and Sex Differences in Wages and Productivity Within and Between Plants." Industrial Relations, Vol. 37, No. 2, April, pp. 232-60.

Hellerstein, Judith K., and David Neumark. 1999. "Sex, Wages, and Productivity: An Empirical Analysis of Israeli Firm-Level Data." International Economic Review, Vol. 40, No. 1, February, pp. 95-124.

Hellerstein, Judith K., and David Neumark. 2003. "Ethnicity, Language, and Workplace Segregation: Evidence from a New Matched Employer-Employee Data Set." Annales D'Économie et de Statistique, 71-72, July-December, pp. 19-78.

- Hellerstein, Judith K., and David Neumark. 2004. "Production Function and Wage Equation Estimation with Heterogeneous Labor: Evidence from a New Matched Employer-Employee Data Set." NBER Working Paper No. 10325.
- Hellerstein, Judith K., David Neumark, and Kenneth R. Troske. 1999. "Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production Functions and Wage Equations." Journal of Labor Economics, Vol. 17, No. 3, pp. 409-46.
- Hellerstein, Judith K., David Neumark, and Kenneth R. Troske. 2002. "Market Forces and Sex Discrimination." Journal of Human Resources, Vol. 37, No. 3, Spring, pp. 353-80.
- Holzer, Harry J. 1990. "The Determinants of Employee Productivity and Earnings." Industrial Relations, Vol. 29, Fall, pp. 403-22.
- Kahn, Shulamit. 1995. "Women in the Economics Profession." Journal of Economic Perspectives, Vol. 9, No. 4, Fall, pp. 193-206.
- Kawaguchi, Daiji. 2003. "A Market Test for Sex Discrimination: Evidence from Japanese Firm-Level Panel Data." Unpublished paper, Institute of Policy and Planning Services, University of Tsukuba.
- Levine, David, and Jonathon Leonard. 2002. "Diversity, Discrimination, and Performance." Unpublished paper, University of California, Berkeley.
- Light, Audrey, and Manuelita Ureta. 1995. "Early-Career Work Experience and Gender Wage Differentials." Journal of Labor Economics, Vol. 13, No. 1, January, pp. 121-54.
- Macpherson, David A., and Barry T. Hirsch. 1995. "Wages and Gender Composition: Why Do Women's Jobs Pay Less?" Journal of Labor Economics, Vol. 13, No. 3, July, pp. 426-71.
- McGuckin, Robert, and Pascoe, George. 1988. "The Longitudinal Research Database (LRD): Status and Research Possibilities." Survey of Current Business, November, pp. 30-7.
- Medoff, James L., and Katherine G. Abraham. 1980. "Experience, Performance, and Earnings." Quarterly Journal of Economics, Vol. 95, No. 4, December, pp. 703-36.
- Mincer, Jacob, and Solomon Polachek. 1974. "Family Investments in Human Capital: Earnings of Women." In T.W. Schultz, ed., Economics of the Family (Chicago: University of Chicago Press), pp. 397-429.
- Neal, Derek A., and William R. Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences." Journal of Political Economy, Vol. 104, No. 5, October, pp. 869-95.
- Neumark, David. 1988. "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination." Journal of Human Resources, Vol. 23, No. 3, Summer, pp. 279-95.
- Neumark, David. 1996. "Sex Discrimination in Restaurant Hiring: An Audit Study." Quarterly Journal of Economics, Vol. 111, No. 3, August, pp. 915-41.
- Neumark, David. 1999. "Wage Differentials by Race and Sex: The Roles of Taste Discrimination and Labor Market Information." Industrial Relations, Vol. 38, No. 3, July, pp. 414-45.

Oaxaca, Ronald L. 1973. "Male-Female Differentials in Urban Labor Markets." International Economic Review, Vol. 14, No. 3, October, pp. 693-709.

Oaxaca, Ronald L, and Michael R. Ransom. 1994. "On Discrimination and the Decomposition of Wage Differentials." Journal of Econometrics, Vol. 61, No. 1, March, pp. 5-21.

O'Neill, June. 1994. "Discrimination and Income Differences." In Susan F. Feiner, ed., Race and Gender in the American Economy (Englewood Cliffs, NJ: Prentice-Hall, Inc.), pp. 13-7.

Troske, Kenneth R. 1998. "The Worker-Establishment Characteristics Database." In John Haltiwanger, Marilyn Manser, and Robert Topel, eds., Labor Statistics Measurement Issues (Chicago: University of Chicago Press), pp. 371-404.

Turner, Margery Austin, Michael Fix, and Raymond J. Struyk. 1991. "Opportunities Denied, Opportunities Diminished: Racial Discrimination in Hiring." Urban Institute Report 91-9, Washington, DC, The Urban Institute.

Table 1: Results from Production Function-Based Tests of Wage Discrimination by Sex

	$\varphi_F$	$\lambda_F$	p-value, $\varphi_F = \lambda_F$
Translog output production function	.84	.55	.00
Translog value-added production function	.83	.56	.00
Translog output production function, IV for materials with lagged materials	1.01	.55	.00
Translog output production function, dropping dimensionality restrictions on sex and occupation			
Production	.71	.49	.00
Managerial/professional	.89	.70	.33
Technical, sales, administrative, and service	1.29	.70	.01
Translog output production function, production and nonproduction workers as imperfect substitutes	.90	.56	.00

Estimates are from Hellerstein, et al. (1999). All estimates are nonlinear least squares, except where otherwise noted.