

NBER WORKING PAPER SERIES

WAGES, PRODUCTIVITY, AND
WORKER CHARACTERISTICS: EVIDENCE
FROM PLANT-LEVEL PRODUCTION
FUNCTIONS AND WAGE EQUATIONS

Judith K. Hellerstein
David Neumark
Kenneth R. Troske

Working Paper 5626

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1996

We are grateful to Joe Altonji, Zvi Griliches, Bruce Meyer, Dek Terrell, and seminar participants at the University of Chicago, Hebrew University, Michigan, Michigan State, Northwestern, NYU, UCLA, UCSD, the FTC, the Census Bureau, and the NBER for helpful comments, and to Daniel Hansen for research assistance. The opinions expressed herein are solely those of the authors and do not in any way reflect the views of the U.S. Census Bureau or the National Bureau of Economic Research. This research was supported by NSF grant SBR95-10876. Neumark's research was also supported by NIA grant K01-AG00589. This paper is part of NBER's research program in Labor Studies.

© 1996 by Judith K. Hellerstein, David Neumark and Kenneth R. Troske. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

WAGES, PRODUCTIVITY, AND
WORKER CHARACTERISTICS: EVIDENCE
FROM PLANT-LEVEL PRODUCTION
FUNCTIONS AND WAGE EQUATIONS

ABSTRACT

We use a unique new data set that combines individual worker data with data on workers' employers to estimate plant-level production functions and wage equations, and thus to compare relative marginal products and relative wages for various groups of workers. The data and empirical framework lead to new evidence on numerous questions regarding the determination of wages, questions that hinge on the relationship between wages and marginal products of workers in different demographic groups. These include race and sex discrimination in wages, the causes of rising wages over the life cycle, and the returns to marriage. First, workers who have ever been married are more productive than never-married workers and are paid accordingly. Second, prime-aged workers (aged 35-54) are equally as productive as younger workers, and in some specifications are estimated to receive higher wages. However, older workers (aged 55+) are less productive than younger workers but are paid more. Third, the data indicate no difference between the relative wage and relative productivity of black workers. Finally, with the exception of managerial and professional occupations, women are paid about 25-35% less than men, but estimated productivity differentials for women are generally no larger than 15%, and significantly smaller than the pay differential.

Judith K. Hellerstein
Department of Economics
Northwestern University
2003 Sheridan Road
Evanston, IL 60208
and NBER

David Neumark
Department of Economics
Michigan State University
East Lansing, MI 48824
and NBER

Kenneth R. Troske
Center for Economic Studies
EPCD, Room 211-WPII
U.S. Census Bureau
Washington, DC 20233

I. Introduction

Competing models of wage determination hinge on the relationships between wages, productivity, and worker characteristics. However, direct measures of worker productivity are hard to obtain, so economists usually rely on proxies for productivity when conducting empirical research. The difficulty with this approach is that whether these proxies reflect productivity is always in doubt, making it difficult to distinguish between competing models.

This lack of data on worker productivity plagues numerous areas of empirical research related to issues of wage determination. For example, with data only on wages and worker characteristics over the life cycle, it is difficult to distinguish human capital models of wage growth (such as Ben-Porath, 1967; Becker, 1975; Mincer, 1974) from incentive-compatible models of wage growth (Lazear, 1979) or forced-saving models of life-cycle wage profiles (Loewenstein and Sicherman, 1991; Frank and Hutchens, 1993). Typical wage regression results report positive coefficients on age, conditional on a variety of covariates, but these positive coefficients neither imply that older workers are more productive than younger ones, nor that wages rise faster than productivity. Similarly, without direct measures of the relative productivities of workers, discrimination by sex, race, or marital status cannot be established based on significant negative estimated coefficients on female or black dummy variables, or positive estimated coefficients on a married dummy variable, in standard wage regressions, since the usual individual-level wage regression controls may not fully capture productivity differences (e.g., Becker, 1985).

To overcome these difficulties, we use a unique new data set that combines individual worker data with data on workers' employers to estimate and compare relative marginal products and relative wages for various groups of workers. This employer-employee data set, the Worker Establishment Characteristics Database (WECD), matches long-form respondents to the 1990 Decennial Census to data on their employers from the Longitudinal Research Database (LRD). These data are a major

improvement over previously available data sources because they combine detailed demographic information on workers in a sample of plants with information on plant-level inputs, outputs, and labor costs.¹

We use these data to estimate production functions in which workers with different characteristics are perfectly substitutable labor inputs with potentially different marginal products, and plant-level earnings equations which represent the aggregation of individual-level regressions over workers employed in a plant. The estimates of these equations allow us to compare the relative marginal productivities and relative wages of workers distinguished by various demographic characteristics.^{2,3} Thus, the data and empirical framework lead to new evidence on numerous topics regarding the determination of wages, including race and sex discrimination in wages, the causes of rising wages over the life cycle, and the returns to marriage.

II. The Relationship Between Wages and Productivity

II.1 The Null Hypothesis

In order to motivate the approach we take in this paper, we first present the simplest model illustrating the relationship between wages and productivity under perfect competition. Consider an

¹However, they are somewhat limited in that they are only cross-sectional, only cover the manufacturing sector, and are weighted toward large plants.

²The WECD is a very rich and useful data set, and has so far been utilized only in a few other studies (Troske, 1994; Carrington and Troske, 1994). There are clearly many important issues which these data may be able to address; we limit this paper solely to the analysis of the relationship between the productivity and wage differentials among workers with different demographic characteristics.

³This paper builds on the framework used in Hellerstein and Neumark (1994 and 1995, hereafter HNa and HNb) to analyze Israeli manufacturing data (although the WECD offers numerous advantages over the Israeli data), and it represents a departure from most of the existing empirical literature on wage determination. As discussed in HNa and HNb, there is little existing research comparing productivity and wage data, and even less using firm-level data. Brown and Medoff (1978) estimate a production function using state-by-industry level data to test whether the union wage premium is associated with higher productivity of union labor. Leonard (1984) uses similar data over time to examine the impact of affirmative action laws on productivity in the U.S. One firm-level productivity and wage study examines evidence of sex discrimination using data from the nineteenth-century French textile industry (Cox and Nye, 1989). Studies applied to more narrowly-defined industries have been pursued in the union literature (Allen, 1984; Clark, 1980). Other research has used proxies for productivity, including using piece-rate pay to measure productivity in time-rate work (Foster and Rosenzweig, 1993) and performance ratings (Holzer, 1990; Korenman and Neumark, 1991; Medoff and Abraham, 1980).

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

$$(1) \quad Y = F(L_1 + \phi L_2),$$

where ϕ 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 the 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 λ . Given this setup, the proportional mix of the two types of labor in each plant will be determined by the relationship between ϕ and λ . If $\phi = \lambda$, then under profit maximization or cost minimization plants will be indifferent to the proportional mix of the two types of labor in the plant. If there is a wedge between the relative marginal product and relative wage so that $\phi \neq \lambda$, then profit-maximizing or cost-minimizing plants will be at a corner solution, hiring either only workers of type L_1 (if $\phi < \lambda$) or only workers of type L_2 (if $\phi > \lambda$). The only equilibrium in this model is when wages adjust so that $\phi = \lambda$, and plants are indifferent between the two types of labor.

Evidence that $\phi \neq \lambda$ is inconsistent with the assumption that we are observing profit-maximizing or cost-minimizing plants in a competitive spot labor market.⁴ This paper can be interpreted as providing empirical tests of this characterization of labor markets. We estimate variants of the plant-level production function in equation (1) simultaneously with plant-level wage equations in order to obtain estimates of ϕ and λ for various types of workers. We interpret cases where we cannot reject the equality of ϕ and λ as evidence consistent with competitive spot labor markets. Cases in which we reject the equality of ϕ and λ indicate some deviation from this characterization of labor markets, such as long-term incentive contracts or discrimination.

⁴Labor supply could be less than completely inelastic; as long as market wages remain above reservation wages, the conclusions are unchanged.

II.2 An Alternative Hypothesis: Discrimination

One such deviation that receives a lot of attention in this paper is labor market discrimination. If there is employer discrimination against L_2 labor, as in Becker (1971), then employers maximize utility defined as

$$(2) \quad U(\pi, L_1, L_2) = F(L_1 + \phi L_2) - w_1 L_1 - w_2 L_2 - d \cdot L_2,$$

where d is the discrimination coefficient capturing an employer's distaste for L_2 . In this case utility maximization implies

$$(3) \quad \phi = \lambda + d/w_1$$

implying that $\phi > \lambda$.

If d varies across employers, this case presents the problem that many firms should be at corner solutions. Faced with the market wage ratio λ , firms will hire only L_2 if $d < \phi MP_1 - w_2$, and only L_1 if $d > \phi MP_1 - w_2$. We do not appear to observe this type of segmentation in hiring. However, this predicted segmentation is a result of the particular specification of employers' discriminatory tastes. An alternative, considered in Neumark (1988), is

$$(4) \quad U(\pi, L_1, L_2) = F(L_1 + \phi L_2) - w_1 L_1 - w_2 L_2 - d \cdot [L_2/L_1].$$

In this case, employers care about the relative level of L_2 , rather than the absolute level. With this utility function, maximization of (4) implies

$$(5) \quad \phi = [w_2 + (d/L_1)]/[w_1 - (d \cdot L_2/L_1^2)].$$

In this case the marginal disutility from an additional unit of L_2 labor is not fixed, but depends on the relative level of L_2 . Thus, even if d varies across employers, employers facing the same λ will hire both L_1 and L_2 labor. Of course, those with a higher value of d will hire less L_2 and more L_1 , as equation (5) shows. Thus, the simple employer discrimination model with heterogeneity in

discriminatory tastes does not preclude all (or most) firms hiring both types of labor, even though they face the same wage ratio.⁵

III. A Structural Production Function Approach

To estimate parameters corresponding to ϕ --the relative marginal productivities of various types of labor--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 .⁶ In logs, this is

$$(6) \quad \ln(Y) = \ln(A) + \alpha \ln(K) + \beta \ln(M) + \gamma \ln g(K, M, QL) + \mu,$$

where $g(K, M, QL)$ is the second-order terms in the production function (Jorgenson, et al., 1973), and μ is an error term.

For each plant in our data set, we have demographic information on a sample of their workforce from the WECD. We assume that in the quality of labor aggregate QL , workers with different demographic characteristics are perfectly substitutable inputs with potentially different marginal products.⁷ For example, assume that workers are distinguished only by sex. Then QL would be defined as

$$(7) \quad QL = L(1 - \phi_F) \frac{F}{L},$$

where L is the total number of workers in the plant, F is the number of women in the plant, and ϕ_F is the marginal productivity of women relative to men. Substituting equation (7) into equation (6), we

⁵Another well-known objection to this model is that employers with discriminatory tastes against a particular group cannot survive in a competitive marketplace (Becker, 1971). However, Goldberg (1982) shows that we can frame the model in terms of nepotism toward type L_1 labor rather than discrimination against L_2 , in which case the results are qualitatively the same, but discrimination (actually, nepotism) will not be competed away.

⁶The results reported in the paper were very similar when a Cobb-Douglas production function was used. The only noteworthy difference is that the evidence consistent with sex discrimination was stronger.

⁷Issues relating to this specification of the labor input are discussed in Rosen (1983). Below, we report some estimates dropping the perfect substitutes assumption.

obtain a production function with which we can estimate ϕ_F , using plant-level data on output, capital and materials inputs, and data on the number of workers and the sex composition of the workforce.

We actually define QL to assume that workers are distinguished not only by sex but also by: race (black and non-black); marital status (ever married); age (divided into three broad categories--under 35, 35-54, 55 and over); education (defined as having attended 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). A firm's workforce can then be fully described by the proportions of workers in each of 192 possible combinations of demographic groups.

To reduce the dimensionality of the problem, for much of our work we impose two 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 otherwise identical non-black women to non-black men. 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. We impose these restrictions due to data limitations. For each establishment, we do not have data on the actual number of workers in each of the 192 possible combinations of demographic characteristics, but instead estimate that number using our sample of workers matched to the plant. It is likely, therefore, that we cannot obtain accurate estimates of the representation of workers in narrowly defined sets of demographic groups. For example, in many plants there are no workers in our sample in some of the demographic groups, even though it is likely that there are, in fact, some

workers in these groups. Our restrictions on QL reduce the number of sample estimates based on small numbers of workers, as well as the number of parameters. The effects of relaxing these restrictions on QL are considered in the empirical results below. To foreshadow the results, relaxing the equiproportionate assumption with regard to the distribution of workers, even in cases in which it is least likely to hold (such as the distribution of men and women across occupations) has relatively minor consequences for the results.

With these assumptions, the quality of labor term in the production function becomes

$$(8) \quad QL = \gamma \ln \left[(L + (\phi_F - 1)F) \left(1 + (\phi_B - 1) \frac{B}{L} \right) \left(1 + (\phi_R - 1) \frac{R}{L} \right) \left(1 + (\phi_G - 1) \frac{G}{L} \right) \right. \\ \left. \left(1 + (\phi_P - 1) \frac{P}{L} + (\phi_O - 1) \frac{O}{L} \right) \left(1 + (\phi_N - 1) \frac{N}{L} + (\phi_S - 1) \frac{S}{L} + (\phi_C - 1) \frac{C}{L} \right) \right],$$

where B is the number of black workers, R is the number of workers ever married, G is the number of workers who have some college education, P is the number of workers in the plant between the ages of 35 and 54, O is the number of workers who are 55 or older, and N , S , and C are the numbers of workers in the second through fourth occupational categories defined above.⁸ Note that the way QL

⁸For example, suppose workers are distinguished by race and sex. Then the unrestricted quality of labor term is

$$QL = L + (\phi_F - 1)WF + (\phi_B - 1)BM + (\phi_F \cdot \phi_B \cdot \phi_{FXB} - 1)BF,$$

where WF is the number of white females, BM the number of black males, and BF the number of black females. The restriction of equal relative marginal productivities implies $\phi_{FXB} = 1$. The equiproportionate distribution restriction implies $BF = B \cdot (F/L)$, $BF = B(1 - (F/L))$, and $WF = F(1 - (B/L))$. Substituting, we obtain

$$QL = L + (\phi_F - 1)F(1 - (B/L)) + (\phi_B - 1)B(1 - (F/L)) + (\phi_F \cdot \phi_B \cdot \phi_{FXB} - 1)B(F/L),$$

which reduces to

$$QL = (L + (\phi_F - 1)F)(1 + (\phi_B - 1)(B/L)),$$

is defined, productivity differentials between groups are indicated when the estimate of the relevant ϕ is significantly different from one (rather than zero). For example, a finding of $\phi_R = 1.3$ would imply that ever-married workers are 30% more productive than never-married workers.⁹

We also allow productivity to vary by size of plant (see Lucas, 1978; Baily, et al., 1992), industry, region, age of plant, and whether or not the plant is part of a multi-plant firm, by adding controls for these plant-level characteristics to the production function.¹⁰

Because materials are likely to be an endogenous input, when we estimate the production function with output as the dependent variable, we instrument for materials with lagged materials.¹¹ If plants differ systematically (i.e., in a persistent manner) in terms of output, and the differences are correlated with materials, then lagged materials is not a valid instrument. However, if the output differences over time are due to uncorrelated period-specific effects, then a lagged value of materials is a valid instrument.

paralleling equation (8).

⁹In the text of the paper, we sometimes report the estimate of ϕ , and whether it is significantly different from one, and sometimes refer to the implied percentage differential ($\phi - 1$), and whether it is statistically significant (i.e., significantly different from zero). The tables report estimates of the ϕ 's.

¹⁰As Griliches and Ringstad (1971) point out, estimates of the first-order terms in the translog production function are not invariant to the units of the data. We therefore de-mean the (log of) capital, materials, and labor quality inputs prior to estimating the production function, so that the coefficients on the productive inputs in the production function are estimated at the mean of the sample. Following Crepon and Mairesse (1993), we de-mean the log quality of labor term, $\ln(QL)$, by first estimating the translog production function without demeaning, constructing plant-level estimates of $\ln(QL)$, and then taking the mean over the sample of the estimated values of $\ln(QL)$. This allows us to measure the returns to scale parameter by adding up the coefficients on the linear terms.

¹¹To instrument for materials, we form the predicted value of log materials, form the nonlinear variables involving materials using this predicted value, and use the latter as instruments (Bowden and Turkington, 1984). We are most worried about the endogeneity of materials, given that materials inputs are the easiest for firms to adjust in the short-run. Nonetheless, it is possible that capital and labor quality are also endogenous. We unfortunately do not have good instruments for these latter two inputs. First, as we discuss below, the capital measure we use in the production function is actually a measure of lagged book value of capital. Second, the data on the demographic composition of workers in a plant is cross-sectional data, so we have no lagged measures of worker quality, nor do we have any other good candidates for instruments. To the extent that these problems affect the coefficients in the wage and productivity equations similarly, our test for the differences between relative wages and productivities should be unaffected. In Section VIII we return to this issue in the context of omitted variable bias in the production function.

We also estimate a value-added version of the production function, using $\ln(Y-M)$ as the dependent variable. Griliches and Ringstad (1971) list numerous justifications for the value-added specification. First, materials may be a particularly endogenous input, and the value-added specification avoids estimating a coefficient on materials. Second, the value-added specification enhances comparability of data across industries and across establishments within industries, when industries or establishments differ in their degree of vertical integration. Third, the value-added specification can be derived from quite polar production function specifications: one in which the elasticity of substitution between materials and value added is infinite (i.e., $Y = f(K, QL) + M$); and one in which this elasticity of substitution is zero (so that materials have to be used in a fixed proportion to output).

IV. Earnings Differentials Among Workers

We have three compensation measures available in our data set: the plant's total annual wage and salary bill; the plant's total annual wage and salary bill plus expenditures on non-wage compensation; and an estimate of the plant's total annual wage and salary bill derived from our sample of workers matched to the establishment. For simplicity, in the following discussion we refer to each of these measures as the plant's total wages. We examine results with each of the compensation measures.

The plant-level wage equation we estimate for most of the results retains the equiproportionate distribution restriction made in defining QL in the production function. We also (again paralleling the production function) restrict the relative wages of workers within a demographic group to be constant across all other demographic groups. Furthermore, we assume that all workers within each unique set of demographic groupings are paid the same amount, up to a multiplicative random error. Then total log wages in a plant can be written as

$$(9) \quad \ln(w) = a' + \ln\left[(L + (\lambda_F - 1)F) \left(1 + (\lambda_B - 1)\frac{B}{L}\right) \left(1 + (\lambda_R - 1)\frac{R}{L}\right) \left(1 + (\lambda_G - 1)\frac{G}{L}\right) \right. \\ \left. \left(1 + (\lambda_P - 1)\frac{P}{L} + (\lambda_O - 1)\frac{O}{L}\right) \left(1 + (\lambda_N - 1)\frac{N}{L} + (\lambda_S - 1)\frac{S}{L} + (\lambda_C - 1)\frac{C}{L}\right)\right] + \varepsilon,$$

where a' is the log wage of the reference group (non-black, never married, male, no college, young, unskilled production worker) and the λ terms represent the relative wage differentials associated with each characteristic.

This plant-level equation can be interpreted as the aggregation over workers in the plant of the individual-level wage equation. To show this, consider a simpler version of the wage equation involving only men and women. The total wage bill in levels implied by equation (9) is

$$(10) \quad w = w_M(L-F) + w_F F,$$

where w_M and w_F are the average wages of men and women. This can be rewritten as

$$w = w_M(L-F) + \lambda_F w_M F = w_M(L + (\lambda_F - 1)F),$$

which in logs is

$$\ln w = a' + \ln(L + (\lambda_F - 1)F),$$

as in equation (9), where $a' = \ln(w_M)$.

Next, consider the individual-level wage equation in levels

$$(11) \quad w_i = w_M M_i + w_F F_i,$$

where M_i and F_i are dummy variables for men and women, respectively. Clearly, the aggregation of this equation over all workers in the plant yields equation (10), from which, as we have shown, the wage equation (9) can be derived.

We interpret equation (9) not as a behavioral equation but simply a definitional one. It assumes that all plants are wage takers in a competitive labor market so that wages do not vary

systematically across plants.¹² In order to relax this assumption somewhat, in the empirical analysis we allow wages to vary systematically with industry, plant size, region, and age of the plant.¹³ In addition, we include as regressors in the wage equation the capital and materials expenditures of the plant. These inputs in the wage equation may account for the possibility that capital and materials are proxies for unobserved ability of workers, possibly because of complementarities between capital and unobserved dimensions of skill (Griliches, 1970), or they may be proxies for other differences across plants that shift wages.

We estimate equation (9) jointly with equation (6). We then compare estimates of the λ 's with the corresponding estimates of the ϕ 's from the production function, and test whether the relative wages of workers with different demographic characteristics are significantly different from their relative marginal products.

V. The Data

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 with information about their employers in the 1989 LRD. Long-form Census respondents report the location of their employer in the prior week, and the type of business or industry in which they work. The Census Bureau then assigns a code for the location of the employer, corresponding to a unique city block for densely populated areas, or corresponding to a unique place for sparsely populated areas. The Census Bureau also classifies workers into industries using Census industry codes so that, in effect, respondents can be assigned to a unique industry-location cell. The Census Bureau also maintains a complete list of all manufacturing establishments operating in the U.S. in a given year, along with location and

¹²As discussed in Section II, this is the correct assumption to make given that we are testing the null hypothesis of competitive spot labor markets.

¹³We also estimate the wage equation and production function for various subsets of the data, in which case wage differentials across workers are not constrained to be equal in all plants.

industry information for these establishments that is similar to the data available for workers. Thus, it is also possible to assign all plants in the U.S. 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. This results in a data set consisting of 199,558 workers matched to 16,144 plants.

To obtain data on a worker's employer, these data must be matched to the 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). The CM is conducted in years ending in a two or a seven, while the ASM is conducted in all other years for a sample of plants. The LRD contains plant data from every CM since 1963 and every ASM since 1971. 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 refer to 1989, we match the worker data to the 1989 plant data in the LRD. Since 1989 is an ASM year, data are only available for a sample of plants. Furthermore, since plant-level capital stock information is only available in Census years, we require all plants to be in the LRD in both 1989 and 1987.¹⁴ Finally, to increase the representativeness of the sample of workers in each plant, we require plants in our data set to have at least 20 employees in 1989 (as reported in the LRD), and at least 5% of their workforce contained in the WECD. Our final sample contains data on 3,102 plants and 129,606 workers.

Summary statistics for plant-level data are given in Table 1. The average plant has 353 employees,

¹⁴Total capital in the plant is measured as the sum of the end-of-year book value of buildings and machinery in 1987. Again, because 1989 is an ASM year, we use materials from 1987 when we instrument for materials in 1989, since in 1987 materials are available for most firms in the LRD as of 1989.

and on average 12% of a plant's workforce is matched to the plant.¹⁵

Troske (1993) concludes that workers are matched to their correct plants--based on the match rate and on high correlations between variables available in the two data sets--with approximately 5% of workers from the Census long-form represented in the WECD. The matching process does not, however, yield a representative sample of workers, as non-black, male, married workers are over-represented in the WECD. Below we discuss some of the implications of this for our empirical results.

VI. Individual-Level Wage Regressions with the WECD Data

Before turning to the results of the jointly estimated plant-level production function and wage equations, we report in Table 2 the results of individual-level wage regressions using the wage data from the WECD. The wage regression results provide a comparison between the WECD data and standard wage regression results reported elsewhere. More importantly, the plant-level wage equation is derived as the aggregation of individual-level wage regressions, as explained above. Thus, comparing results from the plant-level wage equation with those from the individual-level regressions gives us an informal test of this aggregation, and is also useful in assessing other potential biases in the plant-level estimates.

The sets of regressors in the first two columns of Table 2 are among those typically found in wage regressions, and the results are consistent with standard results. The dependent variable in column (1) is the log hourly wage of the individual, defined as annual earnings divided by the product of weeks worked in the year and usual hours worked per week. The dependent variable in the second column is log annual earnings, so the differentials reflect differences in hourly pay as well

¹⁵We have no fewer than two workers per plant. Table 1 also reports the distribution of plants based on number of workers matched.

as hours worked. We focus on results using the annual earnings data because they most closely parallel the earnings measures available in the LRD.

The estimates display features common to numerous other data sets. There is a significant wage gap between men and women, and a smaller but still significant wage gap between blacks and non-blacks.¹⁶ The estimated life-cycle wage profile has the usual quadratic shape. The positive marriage premium (of 16%) parallels that found elsewhere--typically for men. Below, we examine results that let marriage-wage and marriage-productivity differentials vary by sex.

To see how individual-level wage regression estimates are influenced by the addition of plant-level controls in the WECD, column (3) uses the specification in column (2) but adds the plant-level covariates of total employment, capital, materials expenditures, and age of the establishment, as well as individual occupational controls. Adding these variables shrinks the estimated relationship between earnings and education, particularly for highly-educated workers, consistent with these controls being correlated with worker quality. Nevertheless, a significant positive relationship is still present. The estimated coefficients on the other demographic variables (race, sex, age, marital status) do not change much. As in Troske (1994), including the plant's capital stock causes the plant-size wage premium to be negative.

As explained above, to get more reliable estimates of the demographic composition of plants' workforces, in the plant-level estimation we define workers' demographic characteristics more broadly than is typical for individual-level wage equations. In order to provide direct comparability between individual-level wage equations and the plant-level equations we discuss below, column (4) of Table 2 reports the results of the individual-level regression using the more aggregated forms of

¹⁶A race-wage gap of this magnitude (7%) is standard, and suggests that we may be unable to detect significant differences between blacks and non-blacks in plant-level estimates of wage equations (and production functions).

these characteristics that we use in the plant-level regressions. The only major qualitative difference is that the magnitude of the estimated marriage premium almost doubles in size.

Finally, in the plant-level wage and output equations, identification of productivity and wage differentials associated with demographic characteristics of workers comes from covariation across plants in the demographic composition of the workforce, and output and earnings. If we find evidence that, for example, women are paid less than men, the plant-level data do not enable us to determine whether the lower wages of women come from the segregation of women into low-wage plants (e.g., Blau, 1977), or lower pay for women than men within plants. We can, however, add a plant fixed effect to equation (9) to estimate within-plant wage differentials associated with demographic characteristics of workers. Estimates of the specification in column (4), including plant fixed effects, are reported in column (5). Most of the estimated wage differentials (with the exception of that associated with race) are slightly smaller in absolute value. But the differentials remain, and the slight reductions indicate that the wage differentials we are studying are largely within plants.¹⁷ Given this, it seems valid to interpret the plant-level wage equation (9) as the plant-level aggregation of the individual-level wage equations in column (4) of Table 2. In contrast, if the differentials were largely between plants, we could not confidently interpret our estimates as measuring differences between demographic groups. In the absence of measures of productivity for individual workers, we of course cannot test whether estimated productivity differentials also reflect primarily within-plant differentials; we assume this to be the case.

Moreover, as we show below, the results we obtain from estimating the plant-level wage equation resemble closely the individual-level fixed-effects results in Table 2. As we illustrate, some of the differences that do exist are driven by differences in the measurement of wages. Other

¹⁷Groshen (1991) finds a larger role for between-plant wage variation in the male-female wage gap. However, her results are not very comparable to ours: first, she has much finer controls for occupation; second, she studies only five detailed industries, three of which are not in manufacturing.

differences may be due to the fact that there is some systematic variation across plants. Nonetheless, the qualitative results of the fixed-effects estimates in Table 2 nearly always persist in the plant-level results and, in fact, the point estimates for sex and marriage differentials in the plant-level results are virtually identical to those in Table 2. Finally, it is important to note that the differences that we do find between the plant-level and individual-level wage equation estimates do not indicate systematic biases toward zero in the plant-level estimates. This suggests that measurement error in the estimated demographic composition of plants' workforces (which we discuss further below) is not significantly affecting our results.

VII. Plant-Level Production Function and Wage Equation Estimates: Basic Results

The basic results of the joint estimation of the wage and productivity equations are reported in Table 3. In this table, we use LRD wages and salaries to measure earnings. We have more confidence in this plant-level earnings variable than the variable constructed from Census data, because the Census earnings data measure wages and salaries earned by the worker from all jobs, not necessarily the job in the plant to which the worker is matched. In Section VIII, we report results with Census earnings variables, and report a number of other analyses of the sensitivity of the results and of potential sources of bias.

There are two caveats to the results we report that should be noted at the outset. First, although we only have estimates of the percentage of workers in each demographic group, we have treated the various elements of QL for each plant as known for the purposes of estimation. Therefore, there may be biases in both the coefficient estimates and the standard errors which we report here. Note, however, that under the null hypothesis, measurement error in the estimates of the percentages of workers in each demographic group will affect both the productivity and the wage equations similarly, and it is the comparison between corresponding coefficients in the two equations which is of primary interest. In addition, as noted above, the plant-level wage equation estimates

generally do not differ much from the individual-level estimates, suggesting that the plant-level estimates are not biased much by measurement error. Below we discuss results from a simulation of the extent of measurement-error bias, and conclude that our results are robust to this problem.

Second, some of the estimated wage and productivity differentials by occupation and education are surprising. For example, we typically estimate that the marginal product of managerial/professional labor is below that of unskilled production labor. This result may arise because our data set covers production units. Managerial/professional labor may be concentrated at other sites, and the proportion that happens to be located at the production site may have little to do with output.¹⁸

However, we also tend to find that the productivity of skilled blue-collar workers (precision production, craft, and repair) is below--although not significantly--that of unskilled workers, in the output specification. We also typically find that productivity differentials by education substantially exceed wage differentials by education, although we are aware of no theory that predicts this result; it may be in part related to problems in estimating occupational differentials. However, the results reported below were completely insensitive to constraining the productivity and wage differentials by education or occupation to be the same, so whatever problems may plague estimation of the latter differentials, they do not influence our conclusions.

VII.1 Differentials by Sex

For both the output and value-added specifications, the estimates of ϕ_F and λ_F are reported in the first row of Table 3. The point estimate of the productivity differential, ϕ_F , in the output specification is 0.85, which implies that women are 15% less productive than men, a difference that is significantly different from zero.¹⁹ For the value-added specification, the estimate of ϕ_F is 0.96,

¹⁸In principle, we could compare the results from the full sample with those for the subsample of single-plant firms. The latter, however, constitute only 18% of our plants. Given that some firms have multiple plants in the data set, we may not have fully accounted for the correlations in the errors of our equations. However, of the 3,102 establishments in our data set, 1,731 have no matched establishment.

¹⁹When we do not instrument for materials, the estimate of ϕ_F slightly exceeds one (although not significantly).

and is insignificantly different from one.²⁰

In contrast, the estimates of λ_F --the relative wage--are 0.68 and 0.66 in the two specifications and are both significantly less than one. The p-values from tests of the equality of the wage and productivity differentials, reported in columns (3) and (6), indicate that for both the output and value-added specifications, we reject equality. This implies that, on average, women's lower wages do not reflect lower relative marginal products. One interpretation of this finding is that men and women fill the same jobs in plants, and despite being equally productive, women are paid less, consistent with the standard wage discrimination hypothesis. Based on the most conservative estimate of the difference between the sex wage gap and the sex productivity gap (in the output specification), women are estimated to be 15% less productive than men, but are paid 32% less, consistent with just over half of the wage gap being attributable to discrimination.

There are other possible interpretations of this finding, however, some consistent with discrimination and some not. First, the result may arise because women tend to be employed in lower-paying jobs or occupations (at a level of detail finer than the occupational breakdown we use), although these jobs or occupations are not less productive. This could arise because of taste discrimination as in Becker (1971), coupled with legal impediments to paying unequal wages to men and women in the same job or occupation. We do not have the detailed occupation and job-level information with which to address this alternative discrimination hypothesis, nor would we be likely to be able to address it in the production function context even if we had such data, since it seems unlikely that we could obtain reliable estimates of productivity differentials across narrow

However, we would expect there to be upward endogeneity bias in the estimated coefficient of materials, and, given that there is a fairly strong negative correlation between materials and the percent female, the estimate of ϕ_F would also be upward biased.

²⁰We have also estimated the specifications in Table 3 without including controls for capital and materials usage of the plant in the wage equation. Although these variables are significant in the wage equation, excluding them has almost no quantitative effect on the other estimated coefficients.

occupations and jobs. Second, because we are estimating between-plant relationships, another possibility is that women tend to be employed in lower-paying plants, although these plants are not much less productive, on average, than other plants. This is not primarily the case, however, because as column (5) in Table 2 shows, the sex difference in wages in these data arises mostly within plants. At any rate, in either case we would still conclude that there is a sex gap in wages that does not reflect a difference in marginal productivity.

Alternatively, the finding that women's relative wage is below their relative marginal product may arise if plants vary in the degree to which they have implemented labor-saving technological change. If such change is not fully accounted for in the book value of capital (and we would not expect it to be), and if it tends to eliminate production worker jobs which are disproportionately filled by men, then unobserved technological differences could bias upwards our estimate of ϕ_F , because labor is more productive in plants that have implemented relatively more labor-saving technological change.²¹ (At the same time, the estimate of λ_F would not be biased upwards, if wages are set equal to opportunity wages in other plants.) We would only expect this problem not to arise if technological change is capital augmenting, and is correctly captured in capital price deflators. While these desirable circumstances seem unlikely to hold exactly, Baily, et al. (1992) find that results for total factor productivity regressions are the same using book value of capital and a more carefully constructed capital series based on initial capital stocks and annual investment data. Below, we examine this question further.

VII.2 Black/Non-Black Differentials

The second row of Table 3 reports the estimated coefficients in the production function and

²¹For evidence that technological change has reduced the proportion of production worker employment, see Berman, et al. (1994) and Doms, et al. (1994). We need not assume that the technological change saves on production labor. This assumption simply provides a more compelling reason to believe that mismeasured technology or capital is correlated with the percent female.

wage equation on the percentage of blacks in the plant, and the tests of the equality of these coefficients. The relative productivity of blacks, ϕ_B , is estimated to be 1.09 in the output specification and 1.07 in the value-added specification; these estimates are not significantly different from one. The estimated relative wage is 1.06 and is also not significantly different from one. Not surprisingly given these estimates, the tests reported in columns (3) and (6) do not reject equality of relative wages and relative marginal products of black workers.

The plant-level wage results contrast with the individual-level wage regressions reported in Table 2 (and the commonplace finding), according to which the wage differential between blacks and non-blacks, while small, is significant and negative. Our estimates of ϕ_B and λ_B are likely biased upward for two reasons. First, as discussed in Section V, our sample of workers underestimates the number of blacks working in manufacturing. The under-representation of blacks would cause the estimate of λ_B to be biased away from one. Since the estimate of ϕ_B is likely to be similarly biased, however, the test of the equality of the wage and productivity differentials is still valid.²² Second, the fixed-plant-effects estimates of wage differentials between blacks and non-blacks in the WECD are smaller than the cross-section differentials, indicating that within plants, blacks earn less than non-blacks, but that blacks work in slightly higher-paying plants (see also Carrington and Troske, 1994). This suggests that the production function and wage equation estimates--which use across-plant variation--mask lower relative wages paid to blacks. Nonetheless, our results should be biased towards finding no evidence of discrimination only if blacks tend to work in plants that pay relatively higher wages, but in which productivity is not relatively higher. Given that we cannot measure within-plant productivity differentials between workers, all we can conclude is that our results from the between-plant estimates are not consistent with discrimination against black workers. This

²²Strictly speaking, this is true under the null hypothesis that $\phi_B = \lambda_B$, as explained in HNa. Because, white, male, and married workers are overrepresented in the WECD, the same argument may apply to the estimated ϕ 's and λ 's for women and marital status.

conclusion, however, may be premature, as the within-plant productivity differentials between blacks and non-blacks cannot be explored with these data.

VII.3 Estimated Life-Cycle Profiles

As shown in the third row of Table 3, the estimated productivity of workers aged 35-54 (prime-aged workers) relative to younger workers is 0.98 in the output specification and 1.16 in the value-added specification. Neither estimate is significantly different from one, indicating that the results are consistent with a flat productivity profile through this age range. The results from the wage equation indicate that prime-aged workers are paid 8-9% more than younger workers, and although this differential is statistically significant, the p-values from tests of the equality of the wage and productivity differentials are 0.22 and 0.54, and hence we do not reject the equality of the differentials.

In contrast, the results for workers aged 55 and over, shown in the fourth row, provide more compelling evidence that wage growth exceeds productivity growth. In the output specification, the results indicate that older workers are significantly less productive than younger workers, with their relative marginal productivity estimated to be 0.75. In the value-added specification, this estimate is 0.89, not significantly below one. In both specifications, the wages of workers aged 55 and over are significantly higher than the wages of younger workers, with estimates of 1.09 to 1.11. In the output specification we reject equality of the relative marginal products and wages of older workers, with a p-value of effectively zero. In the value-added specification the evidence is weaker, with a p-value of 0.13.²³

²³We also examined results using the proportions of workers in five age groups (<25, 25-34, 35-44, 45-54, and 55 and over). The estimates revealed no significant evidence against a flat productivity profile through age 54, and, as in the earlier results, significantly lower productivity of workers aged 55 and over. On the other hand, the estimates reveal significantly higher wages for all groups of workers aged 25 and over, relative to the youngest group, and little difference among these groups (although the estimates are not inconsistent with a quadratic earnings profile). Finally, the hypothesis tests indicated that the wage premium for all groups of workers aged 25 and over significantly exceeds any productivity differential. Thus, these results parallel those in Table 3, except that the evidence for workers aged 35-54 is somewhat stronger. Note that we are estimating relative wage and marginal productivity profiles; thus, our

VII.4 The Marriage Premium

In Table 3, the estimated wage differential associated with marriage is approximately 26% and is statistically significant. The productivity differential is estimated to be somewhat higher, at 29% or 58% depending on the specification, and both estimates are statistically significant. The p-value for the test of the equality of the wage and productivity marriage premia is 0.76 in the output specification, and 0.22 in the value-added specification. These findings imply that married workers are compensated for being more productive than unmarried workers. The fact that the marriage wage premium reflects an underlying productivity premium suggests that the premium is not attributable to discrimination in favor of married workers, but is instead consistent with married workers being more productive--whether because of selection or a true productivity effect (see Korenman and Neumark, 1991).

VIII. Analysis of Robustness and Potential Biases

We now turn to numerous analyses of the robustness of the results reported in the previous section, and of potential sources of bias. To some extent we focus on sex differentials in relative wages and marginal products, since these results are likely to be most controversial. Accordingly, we also focus on the output specification, which yields a more conservative estimate of the evidence consistent with discrimination against women.²⁴

VIII.1 Alternative Wage Measures

First, in Panel A of Table 4 we report translog output results using the alternative wage measures.²⁵ As shown in the first row of the table, the estimate of ϕ_F is stable, ranging from 0.84 to

results do not necessarily imply that older workers are paid more than their marginal product.

²⁴In general, results for the analyses that follow were qualitatively similar for the value-added specification.

²⁵Annual total wage and salary information reported by the plant in the LRD corresponds very closely to annual total wage and salary information in the plant estimated from the worker data, with similar means and a correlation of 0.98.

0.86. The estimated earnings differential (λ_F) is similar using LRD compensation instead of wages and salaries. Using the Census earnings measure, the estimate of λ_F drops to 0.57. One reason for the lower estimate of λ_F may be that women are less likely to report earnings from more than one job. The results for black workers are consistent across the different wage measures, as are the results for marriage differentials in wages and productivity.

The estimated life-cycle profiles in Table 4 using the alternative earnings measure from the Census indicate that wages rise significantly faster than productivity (which is estimated to have a flat age profile). Using LRD compensation, the p-values for equality of relative marginal products and wages are 0.09 for prime-aged workers, and effectively zero for older workers. Using Census wages and salaries, the p-values are effectively zero for both age groups. The estimated wage growth in the Census may be upward biased, however, if prime-age workers are more likely to report earnings from more than one job. Nonetheless, the combined evidence in Table 4 for life-cycle wage and productivity profiles suggests that wage profiles are steeper than productivity profiles.

VIII.2 Hours-Weighted Estimates

To this point, the percentages of labor input from each demographic group were estimated from the percentages of workers in each demographic group in the sample of workers matched to each plant. We now report results in which we instead compute these percentages based on hours worked, calculated from weeks worked and usual hours worked per week as reported in the Census. As for the Census earnings measure, these data are problematic because hours worked do not necessarily correspond to hours worked solely at the matched establishment. On the other hand, it is desirable to try to incorporate hours information to avoid systematic biases in the estimates of the ϕ 's and the λ 's. For example, if women work fewer hours on average (for a single employer) than do men, then the estimates of ϕ_F and λ_F will be downward biased.

The specification reported in Panel B of Table 4 is the hours-weighted counterpart to that in columns (1)-(3) of Table 3. The estimates of ϕ_F and λ_F both rise by 0.11, to 0.96 and 0.79, respectively. Similarly, the estimates of ϕ_R and λ_R (the parameters for married workers) fall by roughly the same amount, presumably reflecting the greater hours worked by married men. The other estimated ϕ 's and λ 's change by less, but, as we would expect, always change in the same direction for a particular demographic characteristic. The similar changes in the estimates of the corresponding ϕ 's and λ 's confirm our general presumption that mismeasured variables may bias estimates of the ϕ 's and the λ 's, but are much less likely to bias estimates of the differences between them. As a result, the qualitative conclusions in Table 4 are the same as those in Table 3. We still find that the sex gap in wages significantly exceeds the productivity gap, that older workers receive relative wages in excess of their relative marginal products, and that there is no evidence consistent with discrimination against blacks, or in favor of ever-married workers.

VIII.3 Relaxing the Equiproportionate and Equal Relative Productivity Restrictions

The way that labor enters the production function and the wage equation to this point is restrictive in two senses. First, the relative marginal productivities (wages) of two types of workers within one demographic group are restricted to equal the relative marginal productivities (wages) of those types of workers within another demographic group. Second, the proportion of workers defined by one demographic group is restricted to be constant across all other groups.

In Panels A and B of Table 5 we relax certain aspects of these restrictions. To explain how we do this, we discuss the methods and results underlying Panel A of Table 5 in detail, before going on to the other results in the table. In Panel A, we relax both types of restrictions with regard to marriage, race, and sex, based on evidence that the marriage wage premium for men does not carry over to women, and that the race differential is larger for men than for women (e.g., Corcoran and

Duncan, 1979). These race and sex differences in wages suggest that it is of interest to relax the restriction of equal relative marginal products that we imposed on QL , to see if similar patterns are detectable with respect to productivity differentials. In the production function, this yields a quality of labor term of the form:

$$(12) \quad \begin{aligned} QL = & (L + (\phi_F - 1)WFS + (\phi_R - 1)WMR + (\phi_B - 1)BMS + (\phi_R \cdot \phi_B - 1)BMR \\ & + (\phi_R \cdot \phi_F \cdot \phi_{FXR} - 1)WFR + (\phi_F \cdot \phi_B \cdot \phi_{FXB} - 1)BFS + (\phi_F \cdot \phi_B \cdot \phi_R \cdot \phi_{FXR} \cdot \phi_{FXB} - 1)BFR) \\ & (1 + (\phi_G - 1)\frac{G}{L})(1 + (\phi_P - 1)\frac{P}{L} + (\phi_O - 1)\frac{O}{L})(1 + (\phi_N - 1)\frac{N}{L} + (\phi_S - 1)\frac{S}{L} + (\phi_C - 1)\frac{C}{L}), \end{aligned}$$

where WFS denotes the number of non-black, never-married females in the firm, WMR the number of non-black married males, BMS the number of black, never-married males, etc. Introducing these variables relaxes the equiproportionate restriction regarding the distribution of workers. The term ϕ_{FXR} is the coefficient on the interaction for being female and married. A finding that $\phi_{FXR}=1$ would indicate that the marriage productivity premium for women is no different than that for men. The term ϕ_{FXB} is the interaction coefficient for black females, where $\phi_{FXB}=1$ would imply that the productivity differential between men and women does not vary by race. We similarly augment the wage equation to incorporate these interaction terms. Introducing these parameters relaxes the equal relative productivity restriction.²⁶

We first estimate the unrestricted model, using the expression for QL in equation (12) and the corresponding term in the wage equation. We then test which of the "interaction" coefficients (such as ϕ_{FXR}) are significantly different from one, and reestimate the equations retaining only these; if either the productivity or wage interaction is significant, we retain both of them for the corresponding demographic characteristic. We report the latter (restricted) set of estimates. However, in the table we always use the disaggregated estimates of the distribution of workers (WFS , WMR , etc.).

²⁶One way to see that the formulation in equation (12) is correct is to impose this restriction on the parameters, impose the equiproportionate assumption on the data (e.g., $WMR = (R/L) \cdot (1 - \{F/L\}) \cdot (1 - \{B/L\})$), and note that the original quality of labor term in equation (8) results.

For neither wages nor productivity were any of the estimated interaction coefficients significantly different from one. We therefore report the fully restricted model, but without imposing the equiproportionate restriction on the data.²⁷ The estimates and test results closely parallel the corresponding estimates in Table 3. Thus, the imposition in Table 3 of the equiproportionate assumption on the data--at least for the demographic categories that we have considered here--has little or no effect on the estimates.

In Panel B we carry out a similar exercise, but relax the restrictions with regard to sex and occupation, allowing the proportion of the workforce in each occupation to vary by sex, and wage and productivity differentials to vary by sex across occupations. Our primary interest is in the sensitivity of estimated wage and productivity differentials by sex to these restrictions--especially the restriction that the occupational distribution by sex is the same.

In the unrestricted model, only the interactions for managerial/professional workers and technical, sales, etc., workers were significant, so the specification retaining these is reported in Panel B. In this case, the coefficients for female in the top panel refer to unskilled workers (operators, fabricators, and laborers) and precision production, craft, and repair workers. Among these workers, the relative wage of women is less than their relative marginal productivity (0.64 vs. 0.77), and the estimates are significantly different from each other (p -value = 0.03). The last two rows of the panel report sex differences in wages and productivity for the other two occupational categories.²⁸ The sex gap in wages exceeds that in productivity (significantly) for technical, sales, etc., workers, but for managerial/professional workers, the opposite is true, although the difference between the sex gap in wages and in productivity is insignificant. The extra information that we obtain from Panel B of

²⁷Manipulation of the equations in footnote 8 shows that this leads to a different specification from the fully-restricted version.

²⁸We obtain these by multiplying the parameter estimates for the reference group by the estimated interaction parameters. For example, the marginal productivity of managerial/professional women (relative to managerial/professional men) is $0.77 \times 0.87 = 0.67$.

Table 5, then, is that the evidence consistent with sex discrimination comes from the non-managerial and non-professional occupations, in which 86% of the women in the sample work.

VIII.4 Relaxing the Perfect Substitutes Assumption

We next consider estimates of a production function in which workers are imperfect rather than perfect substitutes. It seems to us most natural to separate labor inputs along occupational lines.

We therefore estimate a production function of the form

$$(13) \quad \ln(Y) = \ln(A) + \alpha \ln(K) + \beta \ln(M) + \gamma_P \ln(QL_P) + \gamma_{NP} \ln(QL_{NP}) + g(K, M, QL_P, QL_{NP}) + \mu,$$

where the subscripts 'P' and 'NP' denote production and non-production workers, respectively, and $g(K, M, QL_P, QL_{NP})$ represents the higher-order terms in the translog production function.²⁹ The QL terms in equation (13) are of the same form as equation (8), but defined for the two subsets of workers. The wage equation is of the same form as equation (9), except that production and non-production workers (in each demographic group) are broken out separately. We assume that the ϕ 's and λ 's are the same for both types of workers, which permits us to focus on the effects of relaxing the perfect substitutes assumption.³⁰ In interpreting these results, recall our earlier caveat that estimates for non-production workers may be problematic because the sample of plants includes production units only. Nonetheless, the evidence indicates that the specification of the labor input in terms of perfect vs. imperfect substitutes does not affect the results; the estimates reported in Panel C of Table 5 lead to very similar conclusions to those for the perfect substitutes specification.³¹

²⁹Production workers include the two blue-collar occupations, and non-production workers include the other two occupations. With this form of the production function, output is zero for any plants without workers in an occupation category. We had to drop 219 plants with either no production workers or no non-production workers in the matched sample of workers. Had we entered all four occupations as imperfectly substitutable labor inputs, we would have had to drop many more observations.

³⁰In estimates relaxing this assumption, the only qualitative difference was that only for production workers (and only for older workers) was there significant evidence that wages rise faster than productivity.

³¹In fact, the estimates also change slightly because the sample is smaller. When we estimated the perfect substitutes output specification on this sample, all of the statistical conclusions were the same. However, the relative marginal

VIII.5 Results for Subsamples of the Data

The WECD only contains information on a cross-section of workers from 1990. Because of this, we are unable to account formally (say, through a fixed-effects analysis) for differences across plants in unobservables that may be correlated with the demographic characteristics of a plant. If these unobservable plant-level characteristics affect both the productivity of the plant and the wages that the plant pays (which would occur if labor markets are segmented along the dimension of these unobservables), then, as we argue above, we would expect the omission of these plant-level characteristics to bias in the same way both our estimates of productivity differences and wage differences so as not to negate our test of the equality of the two differentials. If, however, these plant-level characteristics solely affect the productivity of the plant but do not affect wages, then we have reason to worry that only our estimated productivity differentials are biased. While we cannot account for all unobservable plant-level differences, in this section we try to get a sense of the magnitude of this problem by breaking up the sample along dimensions in which we think plants may differ. By doing this, we account for at least some differences across plants that may be related to their demographic composition. In the next section, we look at additional evidence on the role of unobservables.

Durables and Nondurables

In Panel A of Table 6, we split the sample into those plants producing durable goods and those producing nondurable goods. This division is informative because women tend to work in plants that produce nondurable goods, such as textiles, and because the percent female in a plant is negatively correlated with the use of materials in that plant. These considerations suggest that

product of women was noticeably higher (0.94 vs. 0.85 in Table 3), as was the relative marginal product of married workers (1.50 vs. 1.29 in Table 3).

women may tend to work in industries with production technologies different from those in which men work.

The results for both durable- and nondurable-producing plants are similar to the full-sample results. The estimates of the sex differential in productivity suggest that women may be less productive, especially in nondurables. But the estimated wage gaps (40% in nondurables, and 32% in durables) are both significantly larger than the productivity gaps, consistent with sex discrimination. The results for the other coefficients are consistent with the full-sample results, although the estimates are less precise. The results for the tests of the equality of the productivity and wage differentials are unchanged, with older workers earning relative wages that exceed their relative marginal products.

Percent Female

In Panel B of Table 6, we divide the sample into plants with above- and below-median percentages female in the workforce, for three reasons. First, the nature or extent of sex discrimination may differ in plants with varying proportions of female workers. Second, women may work disproportionately in plants with technologies different from plants which employ mostly men. Finally, we can help address the argument that labor-saving technology biases upward the estimate of ϕ_F , which we return to below. If our previous estimates of ϕ_F were biased upward because of variation across plants in labor-saving technology coupled with positive covariation between such technology and the percent female, then when we split the sample into plants with relatively higher or lower percentages of women, the estimate of ϕ_F should fall, as we effectively condition on this technology.

The results are different for the two subsamples. First, in the predominantly female plants women are estimated to be 21% less productive than men and to be paid 26% percent less. These differentials are not significantly different from each other, consistent with no sex discrimination. In

the predominantly male plants, the estimated productivity differential falls to 1%, although the wage differential stays large at 28%. There is a large standard error on the productivity differential, however, so that even though the gap between the point estimates of the relative productivity and wage is 27%, we cannot formally reject the equality of the two. The results therefore indicate that the extent of sex discrimination is lower in plants with large numbers of women, and that this finding is driven by differences in relative productivity, not in relative wages. With regard to biases from labor-saving technological change, we do not find that the estimate of ϕ_F falls in both subsamples; moreover, it actually rises in plants with fewer women, casting doubt on this interpretation of our findings.³²

Employment Levels

Finally, we disaggregate the plants into those with employment levels above and below the median. These results may provide some indication of possible differences in the extent of discrimination between large and small plants. In addition, because the division of plants by employment levels corresponds closely to the division of plants based on number of workers matched to a plant, these results give us some sense of the effect that measurement error in the percentage of workers in each demographic category may have on our results. That is, we match large numbers of workers to large plants but few workers to small plants, so that measurement error should be less problematic in large plants.

The estimated degree of discrimination against women (measured by the estimate of $\phi_F - \lambda_F$) is smaller in the smaller plants (16% versus 43%). These results suggest that smaller firms are less able to indulge in sex discrimination, which may be in part because they have less market power

³²Splitting up the sample into industries (as opposed to plants) with high and low percentages of women produces results that more strongly point to differences between plants in the degree of sex discrimination. Specifically, in industries with below-median percent female, there is no estimated productivity gap, but the estimated wage gap is 38% and statistically significantly different from the productivity estimate. For industries with above-median percent female, there is no statistically significant difference between the sex wage and productivity gap.

(Becker, 1971). As with the results for the subsamples broken out by percent female, this is driven by differences in relative productivities between men and women rather than differences in relative wages. In addition, in the smaller plants there is now marginally significant evidence that prime-age males (aged 35-54) are paid in excess of their relative marginal product. Overall, though, the other qualitative conclusions are similar for the large and small plants. In addition, we do not find that estimates of the ϕ 's and λ 's are consistently closer to one in the sample of small plants, as measurement error might lead us to expect (in the absence of other differences in parameters between large and small plants).

VIII.6 Biases from Labor-Saving Technical Innovation

We next examine evidence on whether biases from unmeasured labor-saving technical change can explain why we find that the relative pay of women falls short of their relative marginal product. For a subset of industries we have independent information on technological innovation from the Census Bureau's 1988 Survey of Manufacturing Technology (see Doms, et al., 1994). This survey covered over 10,000 establishments in SIC industries 34-38 (which are high-technology industries). Over 350 establishments in the WECD can be matched to establishments in this survey. The matched establishments can be used to ask whether the use of advanced technologies is in fact associated with lower percentages of production workers, and whether, for those technologies that appear to replace production workers, there also tends to be a higher percent female.

The results are reported in Table 7. Panel A first reports the percent production workers and percent female, for plants that did and did not use each of the advanced technologies indicated. Panel B reports summary findings. The first finding is that there is no evidence that these advanced technologies are associated with fewer production workers; Panel B shows that just over half (nine) of the technologies are associated with a lower percentage of production workers. We also find that the percent female tends to be lower, rather than higher, in plants using the advanced technologies.

As Panel B shows, of those technologies associated with a lower percentage of production workers, seven are associated with a lower percent female, and only one is associated with a higher percent female. Moreover, only two of the advanced technologies are associated with a higher percent female. Thus, for this subset of industries at least, there is no evidence suggesting that the estimated relative marginal productivity of women is biased upward because the percent female tends to be higher in plants that have installed technology that saves on male production labor, or on male labor generally.³³

VIII.7 Measurement Error

Although we only have estimates of the percentage of workers in each demographic group in each plant, until now we have treated these percentages as known for the purposes of estimation. In this section, we explore more fully the potential effects of measurement error that arises from estimating these percentages. It should be noted that measurement error in the estimates of the percentage of workers in each demographic group will affect both the wage and productivity equations, and it is the comparison between corresponding coefficients in the two equations that is of primary interest. Nonetheless, to the extent that productivity differentials across workers may be of independent interest, and to the extent that measurement error may under some circumstances bias coefficients differently in the wage and productivity equations, it is an issue that merits consideration.

We have already discussed two types of indirect evidence suggesting that measurement error is not causing large biases in our results: first, the results from the individual-level wage equations in Table 2 do not differ greatly from the plant-level wage equation results in Table 3; and second, the

³³The reader might wonder whether the relatively low estimates of the marginal productivity of older workers in the preceding tables stem from a negative association between the use of advanced technologies and the proportion of workers aged 55 and over, perhaps because declining plants do not update their technologies and do not hire as many new workers as other plants. However, the proportion of workers aged 55 and over was virtually the same for plants that did or did not use each of the technologies listed in Table 7. For 13 of the 17 technologies these proportions were equal (to two digits), and the proportions never differed by more than .02. Thus, our estimates of lower productivity of older workers do not appear to be attributable to the association of less-advanced technologies with high proportions of older workers.

results do not differ greatly when we break the sample into large and small plants. In this section, we quantify the magnitudes of measurement error biases with a Monte Carlo simulation.³⁴

Consider the production function and wage equations given by equations (6), (8), and (9).³⁵ From the data, we know the true values of Y , K , M , and L for each plant. All of the remaining variables (F/L , R/L , etc.) are estimated from the sample of T workers within each plant. We simulate the effects of measurement error by creating a synthetic workforce of L workers for each plant. We do this by creating L/T (rounded to the nearest integer) synthetic workers for each of the T workers in the sample. With this new synthetic workforce of L workers, we sample randomly without replacement T workers, and use this simulated sample to estimate the proportions of workers in each demographic group.³⁶ Finally, we use these simulated estimates of these proportions to jointly estimate the production function and wage specifications in columns (1) and (2) of Table 3, obtaining new estimates of the productivity and wage differentials (the ϕ 's and λ 's) across demographic groups. We repeat this process 1000 times, yielding 1000 different values for each of the ϕ 's and λ 's. This procedure enables us to assess the impact on our results of measurement error in the estimated proportions of workers in each demographic group, by comparing model estimates based on the simulated data to model estimates based on the WECD data, which we treat as true. In other words, we assess the impact of measurement error on the estimated ϕ 's and λ 's by adding sampling error to

³⁴Although we know the number of workers sampled in each plant, we do not implement a formal correction for the measurement error bias that results from sampling error. This correction would require a consistent estimate of the variance of the measurement error, which varies by plant depending on the true proportion of workers in any particular category. (For example, at one extreme, in a plant with no female workers the variance of the measurement error in the proportion female is zero.) In other contexts, measurement-error corrections of this type (with non-homogeneous error variances across observations) result in near-singular covariance matrices, because of a high ratio of error variance to total variance (Cockburn and Griliches, 1987).

³⁵In this section we do not estimate a separate coefficient for blacks versus non-blacks. The simulation requires repeated sampling of workers within a plant, and there are too few blacks in the sample to successfully estimate a race coefficient for many of the simulations.

³⁶For example, suppose in the WECD we have four men and six women matched to a plant ($T=10$), and we know from the LRD that there are 100 workers in the plant ($L=100$). We then create a simulated sample of forty men ($4 \times 100/10$) and sixty women ($6 \times 100/10$). Finally, we sample 10 members of this sample, and obtain a new estimate of the proportion female.

the estimated proportions of workers in each demographic group, and re-estimating these parameters.³⁷

Summary results of the 1000 simulations are reported in columns (3) and (4) of Table 8, and can be compared to the results from Table 3 (repeated in columns (1) and (2)) to assess the magnitude of the biases caused by estimating the demographic proportions. The results indicate measurement error biases that, as expected, pull the estimated coefficients toward one, and are greater in magnitude the farther from one is the true value. For example, the mean estimate of ϕ_F in the simulations is 0.89, and the estimate using the actual data is 0.85; the mean estimate of λ_F in the simulations is 0.75, and the estimate using the actual data is 0.68. These results show that the effect of measurement error is to bias us 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.

We therefore conclude from the simulation that our earlier estimates indicating significant estimated gaps between relative wages and productivity for women and older workers are robust to the measurement error problem. On the other hand, we consistently found evidence of the equality of relative wage and productivity premia for married workers, using the actual data. However, the closeness of the point estimates of these premia, the relatively large estimated standard error of the productivity premium, and the similarity between the distributions of the simulated estimates for the wage and productivity premia in Table 8, lead us to believe that even in the absence of measurement

³⁷Because the estimated proportions of workers in each category that we treat as known for our simulation are overdispersed relative to the true population distribution, the extent of measurement error bias indicated by our simulation method is a lower bound for the magnitude of measurement error in the real data.

error, we would still find no significant difference between the wage and productivity premia for married workers.³⁸

IX. Conclusions

This paper uses individual-level data on workers matched with plant-level data on inputs, output and earnings, to estimate the relative wages and relative marginal products of workers with different demographic characteristics. We use a production function approach to plants' technologies, and a simple model of the wage structure of each plant, coupled with a unique data set matching workers and employers. We interpret the results in the context of numerous issues regarding the determination of wage differences between demographic groups, which have previously been addressed based largely on individual-level wage regressions, in the absence of evidence on productivity.

Our basic results indicate that for certain groups of workers, wage differentials do, in fact, match productivity differentials, while for others they do not. Workers who have ever been married are paid more than never-married workers and the wage premium they receive reflects a corresponding productivity premium. This suggests that the marriage premium does not simply reflect discrimination against unmarried workers, but reflects actual productivity differences. However, our data do not distinguish between the hypothesis that marriage reflects an unobservable variable associated with higher wages, and the hypothesis that marriage makes workers more productive.

We find that prime-aged workers (aged 35-54) are equally as productive as their younger counterparts, but in some specifications, the relative wage is significantly higher than any estimated

³⁸The distribution of the productivity premium bounds that of the wage premium. Below the median, the productivity premium is slightly below the wage premium; above the median, the productivity premium is greater than the wage premium. The magnitude of the estimated standard error on the marriage productivity premium in Table 3 alone suggests that measurement error would have to be reducing the gap between the wage and productivity premia by a factor of seven before we would reject equality of the wage and productivity marriage premia.

relative productivity differentials. Results for older workers (aged 55 and over) are more robust, with the evidence generally indicating a wage premium that exceeds any estimated productivity differential (which is, in fact, sometimes significantly negative). This evidence is probably most consistent with Lazear's (1979) explanation of rising earnings over the life cycle.

We consistently find no evidence consistent with discrimination against blacks in manufacturing, in the full-sample as well as the disaggregated results. In addition, there does not appear to be any productivity differential between blacks and non-blacks which might be attributable to pre-market discrimination or other unobserved characteristics, although we are less confident in our separate estimates of the race gap in wages and the race gap in productivity than in our estimates of the difference between them.

Finally, in nearly all of our specifications and samples we find that women are paid significantly less than men, with the wage differential between men and women generally estimated at about 25-35%, depending on the wage measure we use. The productivity differentials between men and women are less robust. The most conservative translog estimates indicate that women may be as much as 15% less productive than men, although this differential may reflect differences in hours worked. Regardless of this, across of a wide variety of analyses the sex wage gap is significantly larger than the sex productivity gap, consistent with sex discrimination. There are only two exceptions. First, for plants with a large proportion of female workers, we estimate that women are less productive and are paid less, but we do not find a wage gap that is much larger than the productivity gap. Second, we do not find evidence consistent with sex discrimination in wages for managerial/professional workers.

As we discuss above, we cannot fully account for unobserved plant-specific differences that are correlated with the quality of workers. If these omitted plant-specific factors affect productivity but not wages (or vice-versa), our tests of the equality of relative wages and relative productivities

may be biased. We have examined this problem as fully as we can within the context of our cross-sectional data, however, and have not found much evidence that supports this. In particular, our conclusion that the sex wage gap is significantly larger than the sex productivity gap in many plants and for most occupations will only be overturned by compelling evidence from other data that women are systematically working in less-productive (but not lower-paying) plants.

We have interpreted the results presented here in a structural framework, in which the estimates from the production function represent relative marginal products of workers and the estimates from the wage equation represent relative wages. We readily acknowledge that the results can be interpreted in a less structural manner, as comparing average productivity differentials across plants to average wage differentials, and relating these differences to the characteristics of the workers in those plants. Even with this interpretation, this research still teaches us considerably more than what we can learn solely from individual-level wage regressions. Finally, under either interpretation of our estimates, our results suggest many important avenues for future research--using the WECD--on the variation of wages and productivity across workers both within and between establishments.

References

Allen, Steven G. 1984. "Unionized Construction Workers Are More Productive." *Quarterly Journal of Economics* 91(4): 543-69.

Baily, Martin Neil, Charles Hulten, and David Campbell. 1992. "Productivity Dynamics in Manufacturing Plants." *Brookings Papers: Microeconomics*, pp. 187-266.

Becker, Gary S. 1985. "Human Capital, Effort, and the Sexual Division of Labor." *Journal of Labor Economics* 3(1, Supp.): S33-S58.

_____. 1975. Human Capital (Second Edition, Chicago: The University of Chicago Press).

_____. 1971. The Economics of Discrimination (Second Edition, Chicago: The University of Chicago Press).

Ben-Porath, Yoram. 1967. "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy* 75(4, Pt. 1): 352-65.

Bergmann, Barbara R. 1974. "Occupational Segregation, Wages, and Profits When Employers Discriminate by Race and Sex." *Eastern Economic Journal* 1: 1199-1221.

Berman, Eli, John Bound, and Zvi Griliches. 1994. "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufactures." *Quarterly Journal of Economics* 109(2): 367-98.

Blau, Francine D. 1977. Equal Pay in the Office (Lexington, MA: Lexington Books).

Bowden, Roger J., and Darrell A. Turkington. 1984. Instrumental Variables (Cambridge: Cambridge University Press).

Brown, Charles, and James L. Medoff. 1978. "Trade Unions in the Production Process." *Journal of Political Economy* 84: 355-78.

Carrington, William and Kenneth Troske. 1994. "Interfirm Racial Segregation." Mimeograph, Johns Hopkins University.

Clark, Kim. 1980. "The Impact of Unionization on Productivity: A Case Study." *Industrial and Labor Relations Review* 33(4): 451-69.

Cockburn, Ian, and Zvi Griliches. 1987. "Industry Effects and Appropriability Measures in the Stock Market's Valuation of R & D and Patents." NBER Working Paper No. 2465.

Corcoran, Mary, and Greg J. Duncan. 1979. "Work History, Labor Force Attachment, and Earnings Differences Between Races and Sexes." *Journal of Human Resources* 14: 3-20.

- Cox, Donald, and John V. Nye. 1989. "Male-Female Wage Discrimination in Nineteenth-Century France." *Journal of Economic History* XLIX(4): 903-20.
- Crepon, Bruno, and Jacques Mairesse. 1993. "Productivite, Recherche-Developpement et Qualifications." *INSEE-Methodes* 37-38: 181-200.
- Doms, Mark, Timothy Dunne, and Kenneth R. Troske. 1995. "Workers, Wages, and Technology." Mimeograph, Center for Economics Studies, U.S. Census Bureau.
- Foster, Andrew D., and Mark R. Rosenzweig. 1993. "Information, Learning, and Wage Rates in Low-Income Rural Areas." *Journal of Human Resources* 28(4): 759-90.
- Frank, Robert H., and Robert M. Hutchens. 1993. "Wages, Seniority, and the Demand for Rising Consumption Profiles." *Journal of Economic Behavior and Organization* 21(2): 251-76.
- Goldberg, Matthew S. 1982. "Discrimination, Nepotism, and Long-Run Wage Differentials." *Quarterly Journal of Economics* 97(2): 307-19.
- Griliches, Zvi. 1970. "Notes on the Role of Education in Production Functions and Growth Accounting." In W. L. Hansen, Ed., Education, Income, and Human Capital (New York: Columbia University Press), pp. 71-115.
- Griliches, Zvi and Vidar Ringstad. 1971. Economies of Scale in the Form of the Production Function (Amsterdam: North-Holland Publishing Company).
- Groshen, Erica. 1991. "The Structure of the Female/Male Wage Differential: Is it Who You Are, What You Do, or Where You Work?" *The Journal of Human Resources*, 226(3): 457-472.
- Hellerstein, Judith K., and David Neumark. 1995. "Are Earnings Profiles Steeper than Productivity Profiles? Evidence from Israeli Firm-Level Data?" *Journal of Human Resources* 30(1): 89-112.
- _____. 1994. "Sex, Wages, and Productivity: An Empirical Analysis of Israeli Firm-Level Data." Mimeograph, Michigan State University.
- Holzer, Harry J. 1990. "The Determinants of Employee Productivity and Earnings." *Industrial Relations* 29(3): 403-22.
- Jorgenson, Dale W., Laurits R. Christensen, and Lawrence J. Lau. 1973. "Transcendental Logarithmic Production Frontiers." *Review of Economics and Statistics* 55(1): 28-45.
- Korenman, Sanders, and David Neumark. 1991. "Does Marriage Really Make Men More Productive?" *Journal of Human Resources* 26(2): 282-307.
- Lazear, Edward P. 1979. "Why Is There Mandatory Retirement?" *Journal of Political Economy* 87(6): 1261-84.

Leonard, Jonathan S. 1984. "Anti-Discrimination or Reverse Discrimination: The Impact of Title VII, Affirmative Action and Changing Demographics on Productivity." *Journal of Human Resources* 19: 145-74.

Loewenstein, George, and Nachum Sicherman. 1991. "Do Workers Prefer Increasing Wage Profiles?" *Journal of Labor Economics* 9(1): 67-84.

Lucas, Robert E., Jr. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics* 9 (Autumn): 508-23.

McGuckin, Robert and George Pascoe. "The Longitudinal Research Database (LRD): Status and Research Possibilities." *Survey of Current Business* (November 1988): 30-37.

Medoff, James L., and Katherine G. Abraham. 1980. "Experience, Performance, and Earnings." *Quarterly Journal of Economics* 95: 703-736.

Mincer, Jacob. 1974. Schooling, Experience, and Earnings (New York: Columbia University Press).

Neumark, David. 1988. "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination." *Journal of Human Resources* 23: 279-95.

Rosen, Sherwin. 1983. "A Note on Aggregation of Skills and Labor Quality." *Journal of Human Resources* 18(3): 425-31.

Troske, Kenneth R. 1993. "The Worker-Establishment Characteristics Database." Mimeograph, Center for Economic Studies, U.S. Census Bureau.

_____. 1994. "Evidence on the Employer Size-Wage Premia From Worker-Establishment Matched Data." Mimeograph, Center for Economic Studies, U.S. Census Bureau.

Table 1
Descriptive Statistics for Matched Establishments

	<u>Mean</u> (1)	<u>Standard deviation</u> (2)		<u>Mean</u> (3)	<u>Standard deviation</u> (4)
<u>LRD Data</u>			<u>Census of Population Data</u>		
Log output (\$1,000)	10.19	1.33	Log estimated wages and salaries (\$1,000)	8.38	1.17
Log value added (\$1,000)	9.34	1.39	Proportion of LRD employment matched	0.12	0.08
Log capital (\$1,000)	8.82	1.53	Proportion with 2-5 workers matched	0.14	
Log cost of materials (\$1,000)	9.43	1.51	Proportion with 6-10 workers matched	0.18	
Log wages and salaries (\$1,000)	8.40	1.17	Proportion with 11-20 workers matched	0.24	
Log compensation costs (\$1,000)	8.62	1.18	Proportion with 21-40 workers matched	0.21	
Employment	353.0	846.8	Proportion with 41+ workers matched	0.24	
Establishment age:			Proportions:		
1-13 years	0.22		Female	0.30	0.23
14-26 years	0.33		Black	0.07	0.12
27-39 years	0.28		Aged 34 or less	0.39	0.20
40+ years	0.17		Aged 35-54	0.48	0.18
Establishment size:			Aged 55 or more	0.13	0.12
1-75 employees	0.22		Some college	0.36	0.21
76-150 employees	0.25		Ever married	0.84	0.14
151-350 employees	0.29		Managerial/professional workers	0.15	0.15
351+ employees	0.25		Technical, sales, administrative, and service workers	0.20	0.15
Industry:			Precision production, craft, and repair workers	0.20	0.15
Food products and tobacco products	0.16		Operators, fabricators, and laborers	0.45	0.22
Textile mill products, apparel, and leather and leather products	0.07				
Lumber and wood products and furniture fixtures	0.04				
Paper and allied products and printing and publishing	0.16				
Chemicals and petroleum refining	0.10				
Rubber and plastics	0.05				
Stone/clay/glass/concrete	0.04				
Primary metals	0.08				
Fabricated metal products	0.08				
Machinery/computer equipment	0.06				
Electrical/Electronic equipment	0.06				
Transportation equipment	0.06				
Instruments/clocks/optical goods and miscellaneous manufacturing	0.03				
Region:					
Northeast	0.29				
Midwest	0.44				
South	0.23				
West	0.05				
Multiple-establishment unit	0.82				

There are 3,102 establishment-level observations, and 128,460 matched individuals from the Census. The sample is restricted to those establishments with total employment of 20 or more, for which at least 5% of employees are matched.

Table 2
Individual-Level Census Log Wage/Earnings Regressions

Dependent Variable:	Specifications with Usual Individual- and Plant-Level Controls			Specifications with Variables Used in Plant-Level Analysis	Fixed Plant Effects
	Log wage (1)	Log earnings (2)	Log earnings (3)	Log earnings (4)	Log earnings (5)
<i>Individual-level variables:</i>					
Female	-0.29 (0.002)	-0.38 (0.003)	-0.33 (0.003)	-0.35 (0.003)	-0.32 (0.003)
Black	-0.03 (0.004)	-0.07 (0.01)	-0.06 (0.01)	-0.05 (0.01)	-0.08 (0.01)
Age	0.05 (0.001)	0.09 (0.001)	0.07 (0.001)
Age ² x 10 ⁻²	-0.05 (0.001)	-0.09 (0.001)	-0.08 (0.001)
Age 35-54	0.21 (0.003)	0.19 (0.003)
Age 55+	0.20 (0.005)	0.18 (0.004)
Ever married	0.10 (0.003)	0.16 (0.004)	0.14 (0.004)	0.27 (0.004)	0.25 (0.004)
<u>Highest degree attained:</u>					
High-school diploma	0.12 (0.003)	0.17 (0.004)	0.13 (0.004)
Some college/ no degree	0.20 (0.004)	0.26 (0.005)	0.18 (0.004)
A.A. degree	0.24 (0.01)	0.32 (0.01)	0.21 (0.01)
B.A. or B.S. degree	0.46 (0.004)	0.56 (0.01)	0.35 (0.01)
Advanced degree	0.62 (0.01)	0.72 (0.01)	0.46 (0.01)
Some college or higher	0.13 (0.003)	0.11 (0.003)
MSA	0.13 (0.003)	0.13 (0.003)	0.09 (0.003)
<u>Dummy variables included for:</u>					
Region (4)	Yes	Yes	Yes	Yes	...
Occupation (one-digit)	No	No	Yes	No	No
Occupation (4)	No	No	No	Yes	Yes

Table 2 (continued)

Dependent Variable:	Specifications with Usual Individual- and Plant-Level Controls			Specifications with Variables Used in Plant-Level Analysis	Fixed Plant Effects
	Log wage (1)	Log earnings (2)	Log earnings (3)	Log earnings (4)	Log earnings (5)
<i>Plant-level variables:</i>					
Log establishment employment	-0.05 (0.002)
Log capital	0.07 (0.002)	0.06 (0.002)	...
Log materials	0.04 (0.002)	0.04 (0.002)	...
<u>Dummy variables included for:</u>					
Industry (two-digit)	Yes	Yes	Yes	No	...
Industry (13)	No	No	No	Yes	...
Establishment size (4)	No	No	No	Yes	...
Establishment age (4)	No	No	Yes	Yes	...
R ²	0.43	0.43	0.49	0.42	...

Standard errors of the estimates are reported in parentheses. The sample size is 128,460. The sample includes all individuals matched to the establishments used in the analysis in the following tables. Less than high-school diploma is the omitted education category.

Table 3
 Joint Production Function and Wage Equation Estimates: Translog Output and Value-Added Production Functions,
 Using LRD Wages and Salaries

	Output specification			Value-added specification		
	Log(output) (1)	Log(wages and salaries) (2)	p-value, col.(1)=col.(2) (3)	Log(value added) (4)	Log(wages and salaries) (5)	p-value, col.(4)=col.(5) (6)
<u>Demographic characteristics:</u>						
Female	0.85 (0.06)	0.68 (0.02)	0.02	0.96 (0.08)	0.66 (0.02)	0.00
Black	1.09 (0.11)	1.06 (0.04)	0.76	1.07 (0.14)	1.06 (0.04)	0.91
Aged 35-54	0.98 (0.07)	1.08 (0.03)	0.22	1.16 (0.11)	1.09 (0.03)	0.54
Aged 55+	0.75 (0.10)	1.09 (0.04)	0.00	0.89 (0.14)	1.11 (0.04)	0.13
Ever married	1.29 (0.15)	1.25 (0.05)	0.76	1.58 (0.26)	1.27 (0.05)	0.22
<u>Productive inputs:</u>						
Log capital	0.09 (0.01)	0.08 (0.01)		0.25 (0.02)	0.11 (0.01)	
Log capital × log capital	0.06 (0.01)	...		0.04 (0.01)	...	
Log capital × log materials	-0.06 (0.01)	
Log capital × log labor quality	0.01 (0.01)	...		-0.03 (0.02)	...	
Log materials	0.54 (0.01)	0.05 (0.01)		
Log materials × log materials	0.17 (0.01)	
Log materials × log labor quality	-0.10 (0.01)	
Log labor quality	0.44 (0.02)	...		0.82 (0.04)	...	
Log labor quality × log labor quality	0.09 (0.02)	...		0.03 (0.04)	...	
<u>Other worker controls:</u>						
Some college	1.74 (0.14)	1.27 (0.04)		1.55 (0.16)	1.28 (0.04)	
Managerial/ professional	0.80 (0.09)	1.12 (0.04)		0.93 (0.13)	1.08 (0.04)	
Technical, sales, administrative and service	1.04 (0.09)	1.14 (0.04)		1.15 (0.12)	1.13 (0.04)	
Precision production, craft, and repair	0.86 (0.09)	1.03 (0.04)		0.99 (0.12)	1.01 (0.04)	

Standard errors of the estimates are reported in parentheses. The sample size is 3,102. Estimates of the intercept are not reported. Test statistics are from Wald tests. The excluded occupation is operators, fabricators, and laborers. Other control variables included in the production function are: industries (13); size (4 categories); region (4); and establishment part of multi-plant firm. Other control variables included in the wage equation are: industries (13); size (4 categories); region (4); and age of plant (4). These control variables were selected by estimating the production function and wage equation jointly without the demographic controls, and retaining those sets of control variables that were jointly significant at the one-percent level. For the output specification, we instrument with log materials in 1989 with log materials in 1987. The model is estimated with the data transformed so that output is homogeneous of degree S in the inputs, where S is the sum of the coefficients of the linear terms for the production function inputs. For variables that enter linearly, we use deviations from the means. For variables that enter nonlinearly, we first estimate the model using the data in levels, and then take deviations from the means of the nonlinear terms. This two-step procedure is valid because the estimated coefficients of all of the nonlinear terms are invariant to the deviations from the mean transformation.

Table 4
 Joint Production Function and Wage Equation Estimates: Translog Value-Added Production Function,
 Using LRD Compensation Costs and Census Wages and Salaries, and Weighting by Hours,
 Estimated Coefficients of Demographic Characteristics

A. Alternative Labor Cost Measures

	<u>LRD Compensation Costs</u>			<u>Census Wages and Salaries</u>		
	<u>Log(output)</u> (1)	<u>Log(compensation)</u> (2)	<u>p-value,</u> <u>col.(1)=col.(2)</u> (3)	<u>Log(output)</u> (4)	<u>Log(wages and salaries)</u> (5)	<u>p-value,</u> <u>col.(4)=col.(5)</u> (6)
Female	0.84 (0.05)	0.67 (0.02)	0.02	0.86 (0.06)	0.57 (0.01)	0.00
Black	1.09 (0.11)	1.05 (0.04)	0.74	1.09 (0.13)	1.02 (0.04)	0.62
Aged 35-54	0.97 (0.07)	1.11 (0.03)	0.09	0.99 (0.08)	1.35 (0.03)	0.00
Aged 55+	0.76 (0.09)	1.15 (0.05)	0.00	0.72 (0.11)	1.27 (0.04)	0.00
Ever married	1.30 (0.14)	1.29 (0.05)	0.95	1.32 (0.17)	1.45 (0.06)	0.48

B. Weighting by Hours

	<u>Log(output)</u> (1)	<u>Log(wages and salaries)</u> (2)	<u>p-value,</u> <u>col.(1)=col.(2)</u> (3)
Female	0.96 (0.06)	0.79 (0.02)	0.00
Black	1.10 (0.12)	1.10 (0.04)	0.96
Aged 35-54	0.97 (0.07)	1.07 (0.03)	0.17
Aged 55+	0.77 (0.10)	1.14 (0.04)	0.00
Ever married	1.13 (0.12)	1.11 (0.04)	0.92

See footnotes to Table 3 for details.

Table 5
 Joint Production Function and Wage Equation Estimates: Translog Value-Added Production Function,
 Using LRD Wages and Salaries, Dropping Equiproportionate Restrictions on Means and Parameters,
 and Relaxing Perfect Substitutes Assumption, Estimated Coefficients of Demographic Characteristics

	A. Drop restrictions on marriage, sex, and race			B. Drop restrictions on sex and occupation			C. Production and Non-Production Workers as Imperfect Substitutes		
	Log(output) (1)	Log(wages & salaries) (2)	p-value, col.(1)=col.(2) (3)	Log(output) (4)	Log(wages & salaries) (5)	p-value, col.(4)=col.(5) (6)	Log(output) (7)	Log(wages & salaries) (8)	p-value, col.(7)=col.(8) (9)
Female	0.85 (0.06)	0.68 (0.02)	0.03	0.77 (0.06)	0.64 (0.02)	0.03	0.98 (0.07)	0.70 (0.02)	0.00
Black	1.08 (0.12)	1.06 (0.04)	0.87	1.13 (0.12)	1.06 (0.04)	0.56	1.16 (0.13)	1.07 (0.04)	0.47
Aged 35-54	0.97 (0.07)	1.07 (0.03)	0.23	0.97 (0.07)	1.08 (0.03)	0.11	1.03 (0.09)	1.06 (0.03)	0.73
Aged 55+	0.75 (0.10)	1.09 (0.04)	0.00	0.74 (0.09)	1.10 (0.04)	0.00	0.77 (0.10)	1.07 (0.04)	0.00
Ever married	1.30 (0.15)	1.25 (0.05)	0.70	1.31 (0.15)	1.25 (0.05)	0.71	1.41 (0.19)	1.37 (0.07)	0.82
Managerial/professional x Female	0.87 (0.25)	1.26 (0.11)
Technical, sales, administrative and service x Female	1.77 (0.33)	1.21 (0.09)
Sex by occupation: Managerial/professional Technical, etc.	0.67 1.36	0.81 0.77	0.44 0.01

See footnotes to Table 3 for details. In Panels A and B, we dropped the equiproportionate restriction in estimating the proportion of workers in each demographic group, and the corresponding restriction on the parameters, and estimated the unrestricted model. We then re-estimated the model imposing those parameter restrictions that were not rejected at the 5% significance level; if either the productivity or wage interaction was significant, we retained both. These latter estimates are reported. The bottom rows report estimated productivity and wage differentials for pairings of demographic groups for which parameter restrictions were not rejected. In Panel C, the sample size is 2,883. It is slightly smaller than in the previous tables because plants with either no production workers or no non-production workers in the matched sample of workers had to be dropped. In these specifications, the relative wages and marginal productivity of production and non-production are allowed to differ, but separate occupational differentials within the production and non-production groups are not included.

Table 6
 Joint Production Function and Wage Equation Estimates: Translog Value-Added Production Function,
 Using LRD Wages and Salaries, Estimated Coefficients of Demographic Characteristics, Subsamples of the Data Set

<i>A. Nondurables and Durables</i>						
	<u>Nondurables (N=1694)</u>			<u>Durables (N=1408)</u>		
	<u>Log(output)</u> (1)	<u>Log(wages and salaries)</u> (2)	<u>p-value, (1)=(2)</u> (3)	<u>Log (output)</u> (4)	<u>Log(wages and salaries)</u> (5)	<u>p-value, (4)=(5)</u> (6)
Female	0.85 (0.08)	0.60 (0.02)	0.00	0.91 (0.07)	0.68 (0.03)	0.00
Black	0.97 (0.14)	1.03 (0.05)	0.64	1.22 (0.16)	1.08 (0.07)	0.36
Aged 35-54	0.98 (0.11)	1.09 (0.04)	0.29	1.07 (0.09)	1.12 (0.05)	0.63
Aged 55+	0.78 (0.15)	1.03 (0.06)	0.07	0.84 (0.12)	1.25 (0.07)	0.00
Ever married	1.23 (0.19)	1.33 (0.08)	0.61	1.20 (0.16)	1.15 (0.07)	0.76
<i>B. High and Low Percent Female</i>						
	<u>Plants Above Median (25% Percent Female (N=1508))</u>			<u>Plants At or Below Median Percent Female (N=1594)</u>		
	0.79 (0.10)	0.74 (0.03)	0.59	0.99 (0.22)	0.72 (0.07)	0.24
Black	1.27 (0.17)	1.06 (0.06)	0.27	0.74 (0.16)	1.03 (0.06)	0.09
Aged 35-54	0.89 (0.11)	1.05 (0.05)	0.19	1.05 (0.10)	1.07 (0.04)	0.89
Aged 55+	0.83 (0.15)	1.07 (0.07)	0.16	0.65 (0.13)	1.11 (0.05)	0.00
Ever married	1.63 (0.31)	1.29 (0.08)	0.27	1.07 (0.15)	1.18 (0.06)	0.50
<i>C. High and Low Employment</i>						
	<u>Below Median (166) Employment (N=1551)</u>			<u>Above Median Employment (N=1551)</u>		
	0.78 (0.05)	0.62 (0.02)	0.00	1.06 (0.11)	0.63 (0.02)	0.00
Black	1.10 (0.11)	1.06 (0.04)	0.73	0.82 (0.16)	1.06 (0.06)	0.12
Aged 35-54	0.98 (0.07)	1.10 (0.03)	0.08	1.11 (0.17)	1.30 (0.07)	0.23
Aged 55+	0.78 (0.09)	1.13 (0.05)	0.00	0.58 (0.20)	1.24 (0.09)	0.00
Ever married	1.31 (0.14)	1.27 (0.05)	0.78	1.25 (0.34)	1.54 (0.15)	0.37

See footnotes to Table 3 for details.

Table 7
Evidence on Percent Production Workers and Percent Female, Based on Technology Use

	Uses technology		Does not use technology	
	% production	% female	% production	% female
	(1)	(2)	(3)	(4)
<u>A. Individual technologies</u>				
<u>Design:</u>				
<i>Computer-aided engineering</i>	.67	.28	.76	.31
<i>Computer-aided design/computer-aided manufacturing</i>	.67	.25	.70	.31
<i>Computer-aided design procurement</i>	.64	.27	.71	.30
<u>Automated material handling:</u>				
<i>Automatic storage and retrieval system</i>	.67	.24	.70	.30
Automatic guided vehicle system	.70	.23	.69	.30
<u>Fabrication, machining, and assembly:</u>				
Flexible manufacturing cell	.69	.27	.69	.30
<i>Computer numerically controlled</i>	.67	.26	.74	.36
Lasers	.70	.28	.69	.29
Pick and place robots	.73	.31	.67	.28
Other robots	.73	.27	.68	.30
<u>Automated sensor based inspection and/or testing equipment:</u>				
<i>Process materials</i>	.67	.29	.70	.29
<i>Final product</i>	.67	.31	.70	.29
<u>Communications and control:</u>				
<i>LAN for technical data</i>	.67	.27	.71	.31
LAN for factory use	.70	.27	.69	.31
Intercompany computer network	.69	.27	.69	.31
Programmable controllers	.70	.27	.67	.34
<i>Computers on factory floor</i>	.68	.28	.71	.31

B. Summary

Number of technologies for which percent female <u>lower</u> for plants using technology:	14
Number of technologies for which percent female <u>higher</u> for plants using technology:	2
Number of technologies associated with lower percent production workers:	9
Of these, number of technologies for which percent female <u>lower</u> for plants using technology:	7
Of these, number of technologies for which percent female <u>higher</u> for plants using technology:	1

Data come from 1988 Survey of Manufacturing Technology, in the following 2-digit SIC industries: fabricated metal products, industrial machinery and equipment, electronic and other electric equipment, transportation equipment, and instruments and related products. The survey covers 10,526 establishments, 358 of which are in our sample. In Panel A, technologies that are associated with lower proportions of production workers are highlighted with italics.

Table 8: Measurement Error Simulation Results

	Estimated productivity <u>differential</u> (1)	Estimated wage <u>differential</u> (2)	Simulated mean productivity <u>differential</u> (3)	Simulated mean <u>wage differential</u> (4)
Female	0.85 (0.06)	0.68 (0.02)	0.89 (0.04)	0.76 (0.01)
Aged 35-54	0.98 (0.07)	1.08 (0.03)	0.98 (0.05)	1.05 (0.02)
Aged 55+	0.75 (0.10)	1.09 (0.04)	0.82 (0.05)	1.09 (0.02)
Ever married	1.29 (0.15)	1.25 (0.05)	1.16 (0.08)	1.13 (0.03)

The estimated productivity and wage differentials in columns (1) and (2) are from Table 3, column (1) and (2). The standard errors of the estimates, and the standard deviations of the simulated values, are reported in parentheses.