Selection, Investment, and Women's Relative Wages over Time

Author(s): Casey B. Mulligan and Yona Rubinstein

Source: *The Quarterly Journal of Economics*, Aug., 2008, Vol. 123, No. 3 (Aug., 2008), pp. 1061-1110

Published by: Oxford University Press

Stable URL: https://www.jstor.org/stable/25098924

SELECTION, INVESTMENT, AND WOMEN'S RELATIVE WAGES OVER TIME*

CASEY B. MULLIGAN AND YONA RUBINSTEIN

In theory, growing wage inequality within gender should cause women to invest more in their market productivity and should differentially pull able women into the workforce. Our paper uses Heckman's two-step estimator and identification at infinity on repeated Current Population Survey cross sections to calculate relative wage series for women since 1970 that hold constant the composition of skills. We find that selection into the female full-time full-year workforce shifted from negative in the 1970s to positive in the 1990s, and that the majority of the apparent narrowing of the gender wage gap reflects changes in female workforce composition. We find the same types of composition changes by measuring husbands' wages and National Longitudinal Survey IQ data as proxies for unobserved skills. Our findings help to explain why growing wage *equality between* genders coincided with growing *inequality within* gender.

I. INTRODUCTION

The changing time allocation of women has been one of the most dramatic economic and social transformations of the past thirty years. Women work more in the marketplace and less at home than they once did (Aguiar and Hurst 2007). Their time in the marketplace has also been transformed, with growing fractions working in once primarily male occupations. Women's professional achievements and pay have grown substantially, although they have not yet fully caught up with men's. At the same time, within-gender wage inequality has increased (Levy and Murnane 1992; Katz and Autor 1999). Inequality grew over this period not only from an increase in the Mincerian returns to education but also because of growing inequality within groups of workers of similar age and education (Katz and Murphy 1992). Growth of inequality during the 1970s, the 1980s, and the 1990s, appears to have occurred throughout the wage distribution as well as over the

*We appreciate the comments of Josh Angrist, David Autor, Gary Becker, Dror Brenner, Steve Durlauf, Raquel Fernandez, Rachel Friedberg, Oded Galor, Reuben Gronau, Jim Heckman, Aitor Lacuesta, Amalia Miller, Kevin M. Murphy, June O'Neill, John Pepper, Chris Rohlfs, Marcia Schafgans, Ed Vytlacil, Yoram Weiss, two anonymous referees, seminar participants at Brown, Chicago, Columbia, Houston, HUJI, MIT, NBER Labor Studies, NYU, Ohio State, Rice, Tel Aviv, Texas A&M, UConn, Wisconsin, and Yale; the research assistance of Ellerie Weber; and the financial support of the National Science Foundation (grant #0241148) and the Sloan Foundation. David Autor generously provided Current Population Survey (CPS) data and programs. We were initially encouraged by Larry Katz to explore in detail the links between wage inequality within and between genders, for which we thank him.

The Quarterly Journal of Economics, August 2008

life cycle (Juhn, Murphy, and Pierce 1993; Gottschalk and Moffitt 1994). Is it possible that inequality within gender, women's professional achievements, and women's time allocation have a unified economic explanation?

Our paper does not attempt to provide a complete explanation of the aggregate increase in women's labor supply, but it does suggest that at least some women's changing time allocations are responses to growing wage inequality within gender and explain a large part of the measured increase in women's relative wages. In particular, we suppose that growing wage inequality within gender indicates a shift in the demand for human capital in favor of those with relatively large amounts of it. In response, women with less human capital may drop out of the workforce, and those with more human capital may enter. Women, especially the more able ones, may also increase their human capital investment. These responses are observed as an increase in various skill proxies—such as schooling and IQ—of the female workforce relative to the female population as a whole, because it becomes more expensive (in terms of opportunity cost) for high-skill women to remain out of the workforce. To the extent that human capital is unmeasured, this response is also observed as an increase in women's measured wages conditional on their observed characteristics. In summary, wage inequality affects the composition of the female workforce. Working women's relative wage growth reflects in part a changing composition of the female workforce.

The evolution of wage inequality within and between genders is consistent with the proposed economic mechanisms. Figure I shows gender *equality* over the years, measured as the log of the median hourly earnings of women working full time full year (FTFY) as a ratio of the median hourly earnings of men working FTFY. Figure I also graphs the evolution of *inequality* within gender, measured as the log of the 90th percentile divided by the 10th percentile in the cross-sectional wage distribution of men working FTFY. We see that both series were flat until about 1977.[1] Both rose—most rapidly at first—from the late 1970s until the mid-1990s. Afterward, both series grew less than they did in the 1980s. Are the series linked by the composition of the female workforce? Or do their comovements have other explanations? This study presents evidence that women's measured relative wages might

1. See also O'Neill (1985) on the apparent constancy of the gender wage gap prior to 1977.

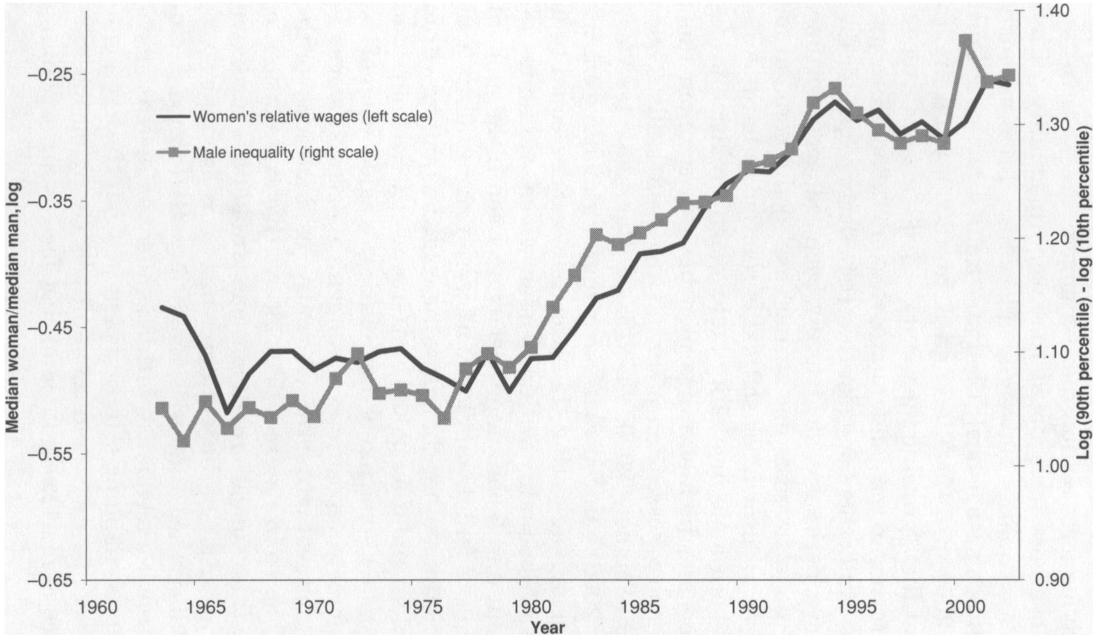

FIGURE I

Wage Inequality within and between Genders

The figure graphs time series of (a) the log of the ratio of the wage of the median working woman to that of the median working man (left scale, no markers), and (b) the log of the ratio of the wage of a man at the 90th percentile of the male wage distribution to that of a man at the 10th percentile (right scale, square markers). The calculations use our CPS wage sample of white persons aged 25–54, without trimming of outliers or adjusting topcodes.

not have grown significantly if the female workforce's composition had been held constant. We also explain how the choice of empirical methods for measuring the effect of composition changes on wages is influenced by the possibility that growing wage inequality was the driving force behind these changes.

Various observers have noted that wage inequality within gender and wage equality between genders have been curiously coincidental, if not paradoxical (Blau and Kahn 1997, p. 2; Card and DiNardo 2002, p. 742). Blau and Kahn (2000, p. 96 [italics added]) suggest that "growing inequality... [is] a major factor *retarding* convergence in the gender gap". Becker (1985a), Katz and Murphy (1992), and others suggest that the effects of gender discrimination have, perhaps coincidentally, been reduced over time. Other explanations suggest that an increase in the demand for "brains" relative to "brawn" or soft skills relative to hard skills, has increased both male inequality and women's relative wages because women and high-earnings men have in common their relatively intensive possession of brains and soft skills (Galor and Weil 1996; Weinberg 2000; Welch 2000; Borghans, ter Weel, and Weinberg 2006).[2] In contrast, our paper highlights women's changing behavior, suggesting that apparent gender equality is a consequence of inequality within gender, which stimulates female investment and labor supply behavior. Even if the amount and distribution of female employment were driven (entirely or in part) by something other than growing wage inequality within gender (see Goldin and Katz [2002] and Greenwood, Seshadri, and Yorukoglu [2005] for some likely possibilities), a calculation of the effect of workforce composition on women's measured wages is still important for evaluating the alternative explanations for the comovements shown in Figure I. (The alternative explanations of Figure I imply that women's relative wages would have increased even if the composition of the workforce had been constant.)

Section II uses the Gronau-Heckman-Roy (GHR) model to illustrate how growing inequality within gender can increase measured wages via a changing selection bias—even if the aggregate female employment rate is held constant—by changing the relative importance of market and nonmarket factors for explaining

2. An alternative explanation is that a relative increase in the demand for women increased their labor supply and relative wages, which in turn depressed the wages of low-skill men (e.g., Topel 1994; Juhn and Kim 1999; Fortin and Lemieux 2000). However, these authors suggest that women are not plausibly better substitutes for low-skill men than for middle- and high-skill men.

which women are employed. Section III shows how the Heckman two-step estimator applied to repeated Current Population Survey (CPS) cross sections suggests that selection into the female workforce shifted from negative in the 1970s to positive in the 1990s, and that the majority of the apparent narrowing of the gender wage gap reflects changes in female workforce composition. Comparisons of various demographic groups also suggest that the selection rule has changed, because demographic groups with high and stable female employment rates have little measured relative wage growth for women, as compared with the significant wage growth measured for other demographic groups. Section IV explains how each of the two selection-correction methods has its own assumptions and data requirements, but nonetheless both come to a common conclusion: although women's wages have grown somewhat—especially when compared with those of men in the left half of the wage distribution—much of the relative wage growth for women shown in Figure I is due to the increased attachment of the most able women to the labor force. Section V explains how "selection" and "investment" have many common economic implications, and how some of our gender-gap selection corrections might be interpreted as adjustments for women's relative human capital investment, but in either case are consistent with the broader conclusion that much of women's relative wage growth is due to their increased supply of human capital to the labor market. Section VI uses husbands' wages from the CPS and IQ data taken from the National Longitudinal Survey (NLS) as proxies for women's unobserved skills to provide nonwage evidence that female workforce selection has shifted from negative to positive, or at least has become less negative over time. Section VII concludes.

II. A Framework for Relating Inequality within Gender to Equality between Genders

We follow much of the previous literature and begin our formal analysis with a log wage equation of the form

$$w_{it} = \mu_t^w + g_i\gamma_t + \sigma_t^w \varepsilon_{it}^w, \tag{1}$$

where w_{it} represents person i's potential log wage in year t, and g_i represents his or her gender (women have $g = 1$, men have $g = 0$). For the moment, we treat μ_t^w as a constant representing determinants of wages that are common to all workers, such as the general levels of supply and demand for human capital.[3] The determinants of wages common to women but not (proportionally) applicable to men, γ_t, may reflect a gap between the human capital of the average woman and the average man, and/or differential market valuation of the average woman's and the average man's human capital.[4] Person i's year t deviation from the average of persons of his or her gender and observed characteristics is $\sigma_t^w \varepsilon_{it}^w$. Following Juhn, Murphy, and Pierce (1991), we have normalized ε_{it}^w so that its standard deviation is 1 and mean 0 for each gender at each point in time.

It follows that γ_t is the average potential log wage for women minus the average potential log wage for men. Estimating time series for $\{\gamma_t\}$ is a necessary component of our argument because, as we show below, it is an indicator of women's wages even when the female workforce composition is held constant. If we could measure potential wages for all men and women regardless of their employment status, the average gender gap measured for each cross section would be sufficient to calculate $\{\gamma_t\}$, because the cross-sectional average for women would be $\gamma_t + \mu_t^w$ and the cross-sectional average for men would be μ_t^w. The potential wage of a person employed FTFY can be measured as her average hourly earnings during the year. For the moment, we interpret the potential wage of a person not employed FTFY as the hourly earnings she would enjoy if she had worked FTFY, but otherwise had the same characteristics. It is well recognized in labor economics that the average wage of working women might not accurately measure the wage of all women, because a number of women do not work and they may not be a random sample of the female population. The gender wage gap G_t among employed persons is calculated by aggregating equation (1) by gender and then subtracting the male average from the female average:

$$G_t = \gamma_t + \sigma_t^w b_t, \tag{2}$$

3. Later we treat μ_t^w as a common function of demographic characteristics, in which case μ_t^w represents wage determinants for a particular demographic group.

4. Forms of wage discrimination common to women are modeled in equation (1) as part of γ_t (Oaxaca 1973). Blau and Kahn (1997) argue that working women have less skill than men, and that the price of skill (regardless of gender) has increased, thereby tending to decrease women's wages relative to men's. If so, this would tend to decrease γ_t (make it more negative) over time. Welch (2000) and Weinberg (2000) argue that men and women have different types of human capital, and the relative price of the two types shifted in favor of women, in effect increasing γ_t over time.

where, for the sake of illustration, equation (2) and subsequent equations ignore the fact that some prime-age white men are also non-employed. Let L_{it} be an indicator for whether person i is employed in year t. The expectation of the idiosyncratic component of wages for employed women is $b_t \equiv E(\varepsilon_{it}^w \mid g_i = 1, L_{it} = 1)$. It differs from zero (its average for the female population) to the extent that non-employed women have different potential wages than employed women. For this reason, b_t is often referred to as a selection bias; it depends on women's behavior in the sense that it is a function of which, and how many, women are employed.

The change in the measured gender gap over time can be represented as

$$\Delta G_t = \Delta \gamma_t + b_{t-1} \Delta \sigma_t^w + \sigma_t^w \Delta b_t. \tag{3}$$

Equation (3) has three terms. The first term is the change in the gender-specific component of net labor demand, which may reflect changes in gender wage discrimination, changes in the market valuation of women's skill endowment (relative to men's), or more rapid accumulation of women's human capital.[5] The second term can also change with the relative market valuation of skill, but in this case the comparison is between the average working woman and the average woman. It does not involve a change in women's behavior in terms of who works and how many work.

The last term is the focus of our paper: it is the change in the standardized selection bias, which, by definition, changes only because women's behavior has changed in terms of the relationship between standardized wages and employment status. Especially during an era when wage inequality has grown significantly (recall Figure I), economic theory makes some suggestions as to the types of behavioral change that might contribute to Δb_t, and thereby the econometric methods that might be appropriate for measuring it. To illustrate this point, we use Roy's (1951) two-sector model as applied by Gronau (1974) and Heckman (1974) to the allocation of women between market and nonmarket sectors.

II.A. The GHR Model for Repeated Cross Sections

The GHR model adds to the potential market wage equation (1) a second "reservation wage," or nonmarket wage equation, in

5. As a measure of net labor demand, the first term would include wage effects of relative supplies occurring because men and women were imperfect substitutes in production.

order to predict which women are employed. Woman i's date t log reservation wage is denoted r_{it}.

$$w_{it} = \mu_t^w + \gamma_t + \sigma_t^w \varepsilon_{it}^w \tag{4a}$$

$$r_{it} = \mu_t^r + \sigma_t^r \varepsilon_{it}^r. \tag{4b}$$

Woman i works at time t if and only if $w_{it} > r_{it}$, which means

$$L_{it} = 1 \quad \text{iff}\ \varepsilon_{it}^r - \frac{\sigma_t^w}{\sigma_t^r}\varepsilon_{it}^w < \frac{\gamma_t + \mu_t^w - \mu_t^r}{\sigma_t^r}. \tag{5}$$

The left-hand side of inequality (5) is person-year specific, whereas the right-hand side is common to all women in a given year. Thus, changes in labor supply have two sources: (i) changes in the parameters affecting only the right-hand side and thereby affecting every woman's employment threshold uniformly, and (ii) changes in (σ^w/σ^r) that affect the selection rule for each woman in a way that depends on her own characteristics.

These two sources of behavioral change have different implications for the validity of various methods for measuring selection bias. Some of these implications are most easily recognized when the error terms ε^w and ε^r follow a standard bivariate normal distribution:[6]

$$\begin{pmatrix} \varepsilon_{it}^w \\ \varepsilon_{it}^r \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \tag{6}$$

where ρ is the cross-sectional correlation between log reservation and potential market wages, assumed to be constant over time. As before, the year t gender gap is a linear function of the bias b_t, but bivariate normality implies a closed-form formula for the bias:

$$G_t = \gamma_t + \sigma_t^w b_t \tag{7}$$

$$b_t \equiv E(\varepsilon_{it}^w \mid g_i = 1, L_{it} = 1)$$

$$= \left[\left(1 + (\sigma_t^w/\sigma_t^r)^2 - 2\rho\sigma_t^w/\sigma_t^r\right)^{-1/2}\left(\sigma_t^w/\sigma_t^r - \rho\right)\right]\lambda_t \tag{8}$$

6. Gronau (1974), Heckman (1979), Keane, Moffit, and Runkle (1988), and Borjas (1994) are among previous studies using the bivariate normality assumption. Heckman and Sedlacek (1985) find a close fit of their extended Roy model (with a joint normality assumption and an instrumental variable) to the U.S. labor market. Moffit (1999) discusses the advantages and disadvantages of the assumption. For our purposes, an important criterion is whether our results can be confirmed by methods without the normality assumptions, or by methods without an instrumental variable.

$$\lambda_t \equiv \phi(\zeta_t)/\Phi(\zeta_t), \quad \zeta_t \equiv \frac{\gamma_t + \mu_t^w - \mu_t^r}{\sigma_t^r}\left(1 + \left(\sigma_t^w/\sigma_t^r\right)^2 - 2\rho\sigma_t^w/\sigma_t^r\right)^{-1/2}, \tag{9}$$

where ϕ and Φ denote the density and cumulative distribution function, respectively, for the standard normal distribution.[7] Comparing formulas (5) and (9) shows that changes in the employment threshold common to all women affects selection bias b in a particular way, through the λ term. The λ term is known as the inverse Mill's ratio, and is a nonnegative and declining function of ζ, which is the common component of the employment threshold transformed to a z-score. In contrast, (σ^w/σ^r) affects the selection rule, and therefore both λ and equation (8)'s square bracket term.

II.B. *Inequality Affects the Selection Rule*

Equation (8)'s square bracket term is the correlation between log market wage w and the net gain $w-r$ from employment (as distinct from the correlation ρ between log market and reservation wages), and is therefore in the [–1,1] interval. If ρ were positive and σ^w were small enough ($\sigma^w < \sigma^r\rho$), then the selection bias would be negative, despite the fact that the inverse Mills ratio is a nonnegative function. In this case, it is said that "selection into the workforce is negative" because the average market wage of employed women is less than the average (potential) market wage of non-employed women. As σ^w increases, the square bracket term becomes positive and approaches 1. This means that an increase in σ^w might cause a fundamental behavioral change, reversing an initial situation in which low-wage women are employed to a situation in which high-wage women are employed, even without

7. To derive equation (8), first substitute the inequality (5) into the definition of b:

$$b_t \equiv E\left(\varepsilon_{it}^w \,\middle|\, \varepsilon_{it}^r - \frac{\sigma_t^w}{\sigma_t^r}\varepsilon_{it}^w < \frac{\gamma_t + \mu_t^w - \mu_t^r}{\sigma_t^r}\right).$$

Bivariate normality implies that ε^w can be decomposed into a linear projection of ε^w on $(\varepsilon^r - \varepsilon^w\sigma^w/\sigma^r)$ plus an orthogonal mean zero error term:

$$\begin{aligned} b_t &= \left[\left(1 + \left(\sigma_t^w/\sigma_t^r\right)^2 - 2\rho\sigma_t^w/\sigma_t^r\right)^{-1/2}\left(\rho - \sigma_t^w/\sigma_t^r\right)\right] \\ &\quad \times E\left(\varepsilon_{it}^r - \frac{\sigma_t^w}{\sigma_t^r}\varepsilon_{it}^w \,\middle|\, \varepsilon_{it}^r - \frac{\sigma_t^w}{\sigma_t^r}\varepsilon_{it}^w < \frac{\gamma_t + \mu_t^w - \mu_t^r}{\sigma_t^r}\right) \end{aligned}$$

where the term in square brackets is the coefficient from a regression of ε^w on $(\varepsilon^r - \varepsilon^w\sigma^w/\sigma^r)$, which is a correlation because ε^w has variance 1. The conditional expectation part of the formula is, by definition, the inverse Mills ratio times -1.

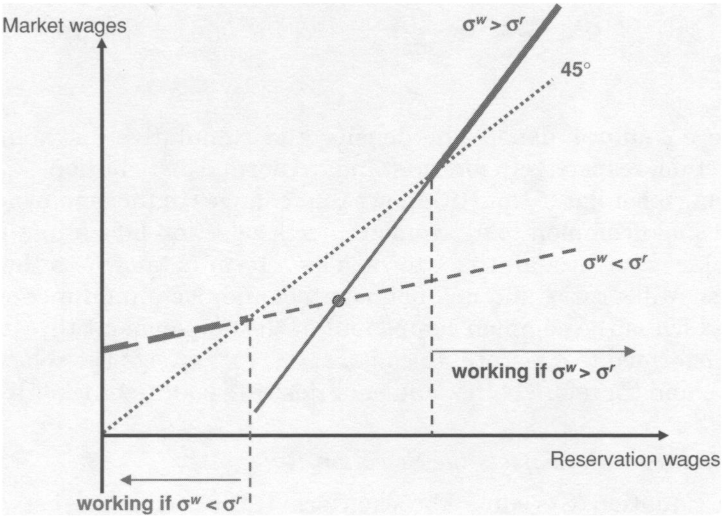

FIGURE II

GHR Model: Inequality Has Composition Effects on Measured Wages

The figure illustrates a comparative static for the Gronau-Heckman-Roy model with respect to σ^w; σ^w (σ^r) denotes the standard deviation of market (reservation) wages. The 45-degree line partitions market workers from nonworkers. Two model parameterizations are shown: σ^w greater (solid line) and σ^r greater (dashed line). The thick portions of the line are above the 45-degree line and indicate workers.

any change in the total amount of female employment. Intuitively, nonwage factors r dominate female employment decisions when σ^w is small, but if σ^w increases enough, market wages can become unequal enough that they dominate nonwage factors as employment determinants, so that nonworking women tend to be the ones with less wage potential.

Figure II illustrates this implication of the GHR model for the special case that productivity in the market and nonmarket sectors is perfectly correlated (i.e., the $\rho = 1$ version of the GHR model). The horizontal axis measures nonmarket productivity r, and the vertical axis measures market productivity w. Any person above (below) the 45-degree line is employed (non-employed). For illustration, r and w are perfectly correlated in the female population, and so all members of the population at a point in time are represented by points on a single straight line whose slope is σ^w/σ^r. If $\sigma^w < \sigma^r$, then the population line crosses the 45-degree line from above and the less-skilled persons work, as shown in Figure II as a dashed line. The measured average wage is the average among working persons, namely those along the dense part of the dashed line. If $\sigma^w > \sigma^r$, then the population line crosses the 45-degree line from below and the more-skilled persons work, as illustrated by the solid line in the figure. In this case, the measured average wage is the average along the dense part of the solid line. The figure shows a dramatic change in measured market wages merely from an increase in σ^w relative to σ^r.[8]

The comparative static formulas with respect σ^w/σ^r and λ are simple in the bivariate normal case:

$$\frac{\partial b_t}{\partial(\sigma_t^w/\sigma_t^r)}\bigg|_{d\lambda_t=0} = \frac{1-\rho^2}{\left(1+(\sigma_t^w/\sigma_t^r)^2 - 2\rho\sigma_t^w/\sigma_t^r\right)^{3/2}}\lambda_t > 0,$$

$$\frac{\partial b_t}{\partial \lambda_t} = \frac{\sigma_t^w/\sigma_t^r - \rho}{\left(1+(\sigma_t^w/\sigma_t^r)^2 - 2\rho\sigma_t^w/\sigma_t^r\right)^{1/2}}. \tag{10}$$

In words, the first comparative static says that changing the selection rule by putting more weight on a woman's rank in the market wage distribution increases the standardized selection bias. If the selection bias is negative, it becomes less negative. If it is positive, it becomes more positive.

The second comparative static can be positive or negative, depending on whether σ^w is greater or less, respectively, than $\sigma^r\rho$. As noted in Section II, σ^w has increased significantly since the 1970s; σ^r also may have increased over time, but if σ^w increased (proportionally) more than σ^r and ρ were constant, then a positive effect of λ on the selection bias is more likely to have occurred in the 1990s than in the 1970s.

If σ^w increased over time enough to change the sign of $\sigma^w - \sigma^r\rho$, then Figure II's dashed line could represent the 1970s' female labor market, and its solid line the 1990s' market. Even if the secular increase in σ^w did not change the sign of the selection-bias formula (8), the square bracket term should be closer to 1 in the 1990s than in the 1970s as long as the effects of increases

8. If the median person is assumed to be in the middle of Figure II horizontally, then median market wage can be calculated geometrically as the vertical position of the (horizontal middle) of the population line, as shown by the circle in the figure. To focus on the σ^w comparative static, in drawing the figure we have held the median market wage constant by having both population lines pass through the circle. As we have drawn it, the circle is below the 45-degree line, and so the majority of persons are not employed in either case. Figure II could have been drawn with alternative assumptions, but, as the algebra below proves, our main conclusion is robust: the composition of the workforce is quite different when $\sigma^w > \rho\sigma^r$ rather than $\sigma^w < \rho\sigma^r$.

in σ^w are not offset by shifts in other parameters. In any case, the fact that any of the parameters are shifting over time means that the selection rule has changed, and ideal empirical methods would be consistent with those changes.

II.C. The Employment Rate and Selection Rule as Separate Determinants of Selection Bias

Equation (9) shows how the inverse Mills ratio λ_t is calculated as $\Phi(\zeta_t)/\Phi(\zeta_t)$, where ζ_t depends on the mean net return to work. Notice that $\Phi(\zeta_t)$ is the fraction P_t of persons who work (i.e., satisfy the inequality (5)). Thus, the inverse Mills ratio λ_t varies over time only to the extent that the employment rate P_t varies. The standardized selection bias can therefore be written as

$$b_t = \theta\left(\sigma_t^w/\sigma_t^r\right)\lambda(P_t), \tag{11}$$

where $\theta(\sigma_t^w/\sigma_t^r)$ is shorthand notation for equation (8)'s square bracket term, which depends on σ_t^w/σ_t^r. The function $\lambda(P)$ is sometimes called a control function (Heckman 2001). Economically, the conditional expectation b is related to the employment rate P because the average worker has a different market wage than the marginal worker, where the marginal worker is added to the workforce by a change in the mean net benefit from working $\gamma_t + \mu_t^w - \mu_t^r$. The gap between the marginal and average worker becomes less relevant as P approaches 1, which is why $\lambda' < 0$ and $\lambda(1) = 0$.

II.D. Empirical Methods Used in Our Paper

The change over time in the behavioral component b_t of the bias in the measured gender gap can be represented as

$$\begin{aligned}\Delta b_t = {} & \theta\left(\sigma_{t-1}^w/\sigma_{t-1}^r\right)[\lambda(P_t) - \lambda(P_{t-1})] + \left[\theta(\sigma_t^w/\sigma_t^r)\right. \\ & \left. - \theta\left(\sigma_{t-1}^w/\sigma_{t-1}^r\right)\right]\lambda(P_t).\end{aligned} \tag{12}$$

The first bracketed term is a change in the selection bias due to a change in the mean net benefit from working, holding fixed the selection rule. The last bracketed term is a change in the selection rule. Either term, or both terms, might explain the changing selection bias in any particular application. We expect that the last bracketed term will be positive because λ is positive (for $P < 1$) and wage inequality has grown over time.

Smith and Ward (1989) suggest that selection bias on the measurement of women's wages had become smaller (less positive) over time during the 1980s as more women entered the labor force. As equation (12) shows, one way to obtain this result is to assume that the selection rule was stable over time and that female workers were "positively selected" (i.e., that $b > 0$ in each year). In this case, all female wages in each year lie on the same downward-sloping control function, with the more recent years having the higher employment rates and less selection bias. However, to the extent that women workers are negatively selected, the first bracketed term actually has the same sign (positive) as the change in labor supply, which means that the selection-bias change (the sum of the two terms) is positive. Even if women workers are positively selected, the last term in equation (12) cannot be neglected because the selection rule in recent years may be different from the rule in earlier years, and the sum of equation (12)'s two terms may well be positive. Furthermore, even in the positive selection case, equation (11) shows that the theory is unambiguous about the change over time in $b_t/\lambda(P_t)$, given the fact that σ^w has increased over time.

Olivetti and Petrongolo (2007) find that countries with higher female employment rates have lower female relative wages. Like Smith and Ward (1989), they interpret their findings as movements along a control function such as the $\lambda(P)$ featured in our model: female wages in countries with higher female employment rates are measured with less selection bias. They also show how the assumption of a fixed selection rule—one that selects the more productive women first—helps to justify quantile approaches to measuring selection bias: with a fixed and positive selection rule, it can be assumed that the nonworking women are always the less productive ones. The number of women not working is known (even though precise values for their productivity are not), and so inferences can be made about higher quantiles of the female wage distribution. However, Olivetti and Petrongolo's quantile methods would not be valid for our purposes if the selection rule were changing over time, so that the fraction of nonworkers who are in the left tail of the wage distribution is also changing.[9] Our Figure II illustrates one (admittedly extreme) situation in which the median potential wage of women is closest to the maximum wage of working women in the 1970s, but closest to the minimum wage of working women in the 1990s.

9. Neal (2004) also emphasizes that female labor force selection can be positive or negative, and that this possibility is important for comparing black women's wages to white women's wages.

Given the background of growing wage inequality within gender, the purpose of our paper is to measure the year effects on the gender wage gap, correcting for selection bias with methods that are consistent with control functions that shift over time. One of those methods, the Heckman (1979) two-step estimator, derives exactly from the GHR model with a bivariate normal distribution. However, alternative methods of measuring selection bias have been proposed in the literature, are consistent with control functions that shift over time, do not rely on a normality assumption, and can be applied here. One alternative is the "identification at infinity" method. Section III applies these methods to repeated cross sections from the CPS.

III. Estimates of Women's Relative Wages from the CPS

III.A. Description of Our CPS Samples

Our work with the Census Bureau's CPS uses the March Annual Demographic Survey files. Our main sample from the surveys is typical of those in the literature: white non-Hispanic adults between the ages of 25 and 54, excluding persons living in group quarters or with missing data on relevant demographics. We classify all persons as either working or not working according to whether they work FTFY (35+ hours per week paid at least 50 weeks during the year). The FTFY part of our main sample is the starting point for our wage sample, which further excludes the self-employed; persons in the military, agricultural, or private household sectors; persons with inconsistent reports on earnings and employment status; and persons with allocated earnings (see Appendix I for additional details, including our treatment of wage topcodes and outliers). We trim wage outliers when calculating mean wages, and keep them when calculating quantiles (as in Figure I).[10] We classify the adult population into six educational categories. Wages are measured as annual earnings divided by

10. Figure I's unmarked solid series is essentially identical to the gender wage gaps reported in the literature. It shows a median log wage gap of –0.47 in the early 1970s, which corresponds to a value of 0.62 for the ratio of median female wage to median male wage and coincides with the longstanding wage ratio of 0.60 discussed by Fuchs (1971, p. 9). Blau and Kahn (2000, Figure 1) report a value of about 0.62 prior to 1975 for the ratio of (arithmetic) average female weekly wage to average male weekly wage (average and median gender gaps are quite similar—compare our results below with Figure I). By the late 1990s, our Figure I shows how the raw gender wage gap closed by 0.18 log points, which corresponds to an increase in the female-male wage ratio by a factor of 1.20. Blau and Kahn (2000, Figure 1), Welch (2000, Figure 3), and many others have calculated time

annual hours. The details of our hours and education coding are also described in Appendix I. We weight all CPS calculations using the March supplement weight.

III.B. Estimates Using Heckman's Two-Step Estimator

In the Heckman two-step model, demographic characteristics are assumed to linearly affect μ_t^w and μ_t^r, but not affect ρ, σ_t^w, or σ_t^r. In particular, $\mathbf{X}$ is a row vector of demographic characteristics affecting market wages (and polynomials thereof), and $\mathbf{Z}$ is the row vector $\mathbf{X}$ plus a vector of additional demographic characteristics affecting only reservation wages. In addition, we assume that selection bias is zero for men. For the purposes of estimation, these assumptions imply that inequality (5) becomes a probit equation (13) for the female employment rate $P_t(\mathbf{Z})$ by demographic group and year and a log market wage equation (14) for employed persons:[11]

$$P_t(\mathbf{Z}) \equiv \text{Prob}(L = 1 \mid \mathbf{Z}, g = 1) = \Phi(\mathbf{Z}\,\delta_t) \tag{13}$$

$$w_{it} = \mathbf{X}_{it}\beta_t + g_i\gamma_t + g_i\theta_t\lambda(\mathbf{Z}_{it}\delta_t) + u_{it} \tag{14}$$

The vector $\mathbf{X}$ includes educational attainment dummies, marital status, a potential work experience quartic interacted with education dummies, and region. The vector $\mathbf{Z}$ has the same elements, plus the number of children aged 0–6 interacted with marital status; β and δ are coefficient vectors. The error term u_{it} is the unobserved component of wages $\sigma_t^w\varepsilon_{it}^w$ from equation (1) minus the inverse Mills ratio term $\theta_t\lambda$.[12]

Following Gronau (1974) and Heckman (1979), our estimation proceeds in two steps (hereafter, the Heckman two-step

series of female-male wage ratios that increased by a factor of 1.2 or 1.25 over that period.

Our calculation of wage inequality among men (Figure I's square-marked series) also accords with calculations in the literature. Our log 90–10 wage differential rises from 1.05 in 1967 to 1.34 in 2002, with most of the increase occurring fairly linearly during the years 1976–1994. Autor, Katz, and Kearney (2005) report a log 90–10 wage differential of 1.10 in 1967 (see also Katz and Autor [1999]), as compared to a value of 1.60 for 2002, with most of the increase occurring fairly linearly between 1975 and 1993. Our calculations would coincide more exactly with theirs if we had used the 18–65 age range, as they did.

11. The inequality (5) describing employed persons is $\varepsilon_{it}^r - \frac{\sigma_t^w}{\sigma_t^r}\varepsilon_{it}^w < \frac{\gamma_t + \mu_t^w - \mu_t^r}{\sigma_t^r}$. The bivariate normality assumption implies that the left-hand side is normally distributed. Linearity of μ_t^w and μ_t^r in $\mathbf{Z}$ implies that the probit index is a linear function of $\mathbf{Z}$. Thus the employment rate is $P_t(\mathbf{Z}) = \Phi(\mathbf{Z}\delta)$. See also Moffitt (1999).

12. By definition, the expected value of u among workers is zero if the model parameters are consistently estimated.

TABLE I
CORRECTING THE GENDER WAGE GAP USING THE HECKMAN TWO-STEP ESTIMATOR

	Method		
Period	OLS	Two-Step	Bias
	Panel A: Variable Weights		
1975–1979	−0.414	−0.337	−0.077
	(0.003)	(0.014)	(0.015)
1995–1999	−0.254	−0.339	0.085
	(0.003)	(0.014)	(0.015)
Change	0.160	−0.002	0.162
	(0.005)	(0.020)	(0.021)
	Panel B: Fixed Weights		
1975–1979	−0.404	−0.330	−0.075
	(0.003)	(0.014)	(0.014)
1995–1999	−0.264	−0.353	0.089
	(0.004)	(0.015)	(0.016)
Change	0.140	−0.024	0.164
	(0.005)	(0.021)	(0.021)

Notes. Each table entry summarizes regression results (reported in full in Appendix II). The entries are female minus male log wages, which differ from each other in terms of (a) rows, i.e., time period used for estimation (1975–1979 vs. 1995–1999); (b) columns, i.e., whether the regression includes the inverse Mills ratio (OLS does not include it, two-step does); and (c) panels, i.e., the weighting used to average the regression results across demographic groups (variable vs. fixed weights). The "Bias" column is the difference between the OLS and two-step columns. The "change" row is the difference between the 1995–1999 and 1975–1979 rows. Weights are factions of working women in each demographic group and are time-specific (variable) or pool both time periods (fixed).

The regressions control for demographics interacted with gender and use our CPS wage sample of white persons aged 25–54, trimming outliers and adjusting topcodes as described in Appendix I.

Bootstrap standard errors are in parentheses.

estimator), separately for every cross section. First, we estimate $P_t(\mathbf{Z})$ as the fitted values from the probit equation above, estimated on a CPS sample of all prime-age white women. The dependent variable for the probit is working FTFY. $P_t(\mathbf{Z})$ is set to 1 for men. Second, for a sample of persons employed FTFY, the log wage equation (14) is estimated using least squares, with a value for the inverse Mill's ratio assigned to each person according to estimates from the probit equation.[13] Tables I and II display the results from the CPS data (information about data processing and the samples used in the CPS regressions is provided in Appendix I). The tables are based on four wage regressions, which differ according to the years sampled (1975–1979 vs. 1995–1999) and whether the inverse Mills ratio is included as a regressor.

13. Standard errors are calculated with the nonparametric pairwise bootstrap method (1,000 replications), and thereby account for the facts that estimation occurs in two stages and that the regression equation error terms are heteroscedastic.

TABLE II
GENDER-GAP CHANGES BY MARITAL STATUS AND SCHOOLING

	OLS			Two-Step	Bias
	1975–1979	1995–1999	Change	Change	Change
	Panel A: All				
Conditional on marital status	−0.404	−0.264	0.140	−0.024	0.164
	(0.003)	(0.004)	(0.005)	(0.021)	(0.021)
Not conditional on marital status	−0.431	−0.270	0.160	−0.009	0.169
	(0.003)	(0.004)	(0.005)	(0.017)	(0.018)
	Panel B: By Marital Status				
Currently married	−0.471	−0.311	0.161	−0.026	0.187
	(0.004)	(0.004)	(0.005)	(0.023)	(0.024)
Separated	−0.380	−0.293	0.087	−0.066	0.153
	(0.021)	(0.021)	(0.030)	(0.035)	(0.046)
Widowed	−0.430	−0.252	0.178	0.019	0.159
	(0.025)	(0.042)	(0.049)	(0.053)	(0.072)
Divorced	−0.326	−0.189	0.136	0.019	0.117
	(0.010)	(0.009)	(0.013)	(0.020)	(0.024)
Never married	−0.179	−0.127	0.052	−0.062	0.114
	(0.010)	(0.009)	(0.013)	(0.019)	(0.023)
	Panel C: By Education				
0 to 8 years	−0.378	−0.322	0.056	−0.206	0.262
	(0.035)	(0.091)	(0.098)	(0.103)	(0.142)
High school, not grad.	−0.429	−0.243	0.185	−0.373	0.222
	(0.018)	(0.032)	(0.037)	(0.046)	(0.059)
High school graduates	−0.427	−0.297	0.130	−0.037	0.167
	(0.007)	(0.009)	(0.011)	(0.023)	(0.026)
Some college	−0.409	−0.258	0.151	−0.008	0.159
	(0.010)	(0.010)	(0.014)	(0.024)	(0.028)
College	−0.400	−0.237	0.163	0.012	0.151
	(0.013)	(0.011)	(0.017)	(0.025)	(0.030)
Advanced degrees	−0.276	−0.179	0.096	−0.018	0.115
	(0.023)	(0.017)	(0.028)	(0.032)	(0.043)

Notes. Each table entry summarizes regression results (reported in full in Appendix II). The entries are female minus male log wages and differ from each other in terms of (a) rows, that is, demographic groups; (b) columns, that is, time period used for estimation and whether the regression includes the inverse Mills ratio (OLS does not include it, two-step does); and (c) panels, i.e., the types of demographic groups summarized. Time-invariant female workforce weights are used to average the regression results across demographic subgroups. The "Bias" column is the difference between the OLS and two-step columns.

The regressions control for demographics (which include marital status unless indicated otherwise) interacted with gender, and use our CPS wage sample of white persons aged 25–54, trimming outliers and adjusting topcodes as described in Appendix I.

Bootstrap standard errors are in parentheses.

For reference, the column or columns of the tables display OLS estimates of the gender gap (the inverse Mills ratio is not a regressor). Consistent with Figure I and the results of previous studies, Table I's first entry (Panel A) shows the 1970s' gender wage gap of

−0.414 log point. By the 1990s (second row), this gap had closed to −0.254. The amount closed (third row) was 0.160 log points. The second column of Table I has the same format as the first but reports two-step estimates, namely, results from wage regressions with the same specification as in the OLS section of the table, except that the inverse Mills ratio is included as a regressor. Table I shows that, controlling for selection bias (i.e., adding the inverse Mills ratio as a regressor), the gender wage gap is −0.337 in the 1970s and −0.339 in the 1990s. Controlling for the inverse Mills ratio, the gender wage gap *widened* slightly (0.002 log points).[14]

The third column of Table I is calculated from the first two columns as the difference between OLS and two-step estimates. According to the GHR model, it is the selection bias included in OLS estimates that by definition does not control for selection because the inverse Mills ratio is not among the regressors. The estimated selection bias is negative in the 1970s, and positive in the 1990s, which suggests that the selection rule was substantially different during the two time periods.

Because Table I is based on regressions in which gender interacts with all other characteristics in the wage equation,[15] the results of any one particular wage regression imply a different gender wage gap for each demographic group. Panel A of the table reports an average gap for all demographic groups, weighted by the fraction of the female FTFY workforce in each demographic group. In principle, the average gender wage gap reported in the top "variable weight" panel—it uses one set of weights for 1975–1979 and another set of weights for 1995–1999—can change over time because the average group has a closing gap, or because the population weights shift over time from high-gap groups to low-gap groups. Panel B, the "fixed weight" panel, uses a common set of weights to calculate all entries: the fractions of FTFY women in each demographic group for the ten pooled years 1975–1979 and 1995–1999. The fact that the variable-weighted OLS gender gap closes more than the fixed-weighted gap (0.160 vs. 0.140) means

14. Maximum likelihood estimates (not shown in Table I) suggest about 0.03 log points less selection-corrected relative wage growth for women, mainly because a more negative coefficient on the inverse Mills ratio is estimated for 1975–1979.

15. Recall that we estimate the probit with female data only. We interact **X** with female in the second-step wage regression in order to ensure that we identify the coefficient on the inverse Mills ratio using the children variable rather than functional-form assumptions (i.e., an assumption that the first stage has gender interactions but the second stage does not). In effect, we estimate the wage regressions separately by gender.

that demographic groups with small OLS gender gaps (such as never-married and advanced degrees) have increased their relative size. This is also true for the Heckman two-step estimates (OLS and two-step average gaps are calculated with a common set of weights).

Measuring wages for prime-age white males may also be subject to selection bias, because their FTFY employment rate is significantly less than 100% (e.g., 76% in 1975–1979 and 79% in 1995–1999 in our sample). Growing wage inequality should increase the importance of market wages, rather than nonwage factors, for selecting men into the workforce, and so the amount of selection bias could have changed for them too, thereby offsetting some of the selection-bias growth for women. Ideally, our estimates based on the Heckman two-step estimator would include a shifter of male labor supply. In the absence of such a variable, we have attempted two ad hoc corrections for male selection-bias growth (not reported in Tables I or II or Figure III based on our benchmark specification) in the Heckman framework.[16] These findings suggest that our benchmark two-step estimates of the gender-gap change might, by failing to correct for male selection bias, understate women's relative wage growth by about 0.01 log points, if at all. The identification-at-infinity method shown in Section III.D adjusts for male selection bias in a more rigorous way.

Table II shows some relationships between gender-gap changes, marital status, and schooling. Panel A displays results for the average female FTFY worker. The first row is a condensed version of Table I's Panel B in which the wage and probit regressions include marital status among the regressors. Because marital status has changed over time and may itself react to the wage structure, it is interesting to see results without marital status regressors, as in Panel A's second row. Both rows show a selection-bias change of 0.16–0.17 log points.

16. One correction finds a male selection-bias change of −0.05 (i.e., in the opposite direction as women's bias change) by adding the inverse Mills ratio to our male wage equations, where the inverse Mills ratio is derived from male probit equations with the same explanatory variables as the wage equation. The second compares Table I's calculations (variable weights) with a version of Table I that gives zero weight to demographic groups in which the male FTFY rate is less than 80% (as a result, about half of women receive zero weight, and the male FTFY employment rate is 84% in 1975–1979 and 87% in 1995–1999). With this version, the gender gap widens 0.01 log points more than in Table I, which suggests that either male selection bias increased 0.01 log points (i.e., in the same direction as women's bias change) or that demographic groups with high male FTFY rates are not representative of all demographic groups in terms of gender wage gap changes.

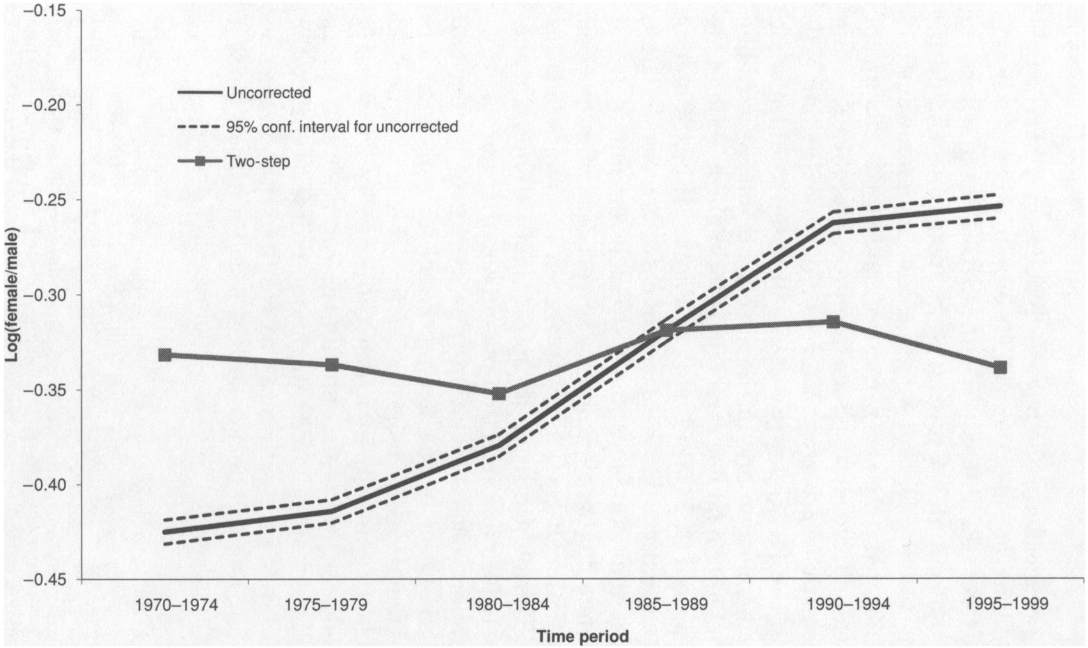

FIGURE III

Correcting the Gender Wage Gap: The Heckman Two-step Estimator

The figure graphs two time series of women's log wages relative to men's (unmarked and square-marked), plus a 95% confidence interval for one of them (dashed). Both relative wage series are net of measured demographic characteristics and averaged across demographic groups using time-specific female workforce weights. Only the two-step series (square-marked) is net of the inverse Mills ratio. The calculations use our CPS sample of white persons aged 25–54, trimming outliers and adjusting topcodes as described in Appendix I.

In comparison with Panel A's first row, Table II's Panels B and C report the partial "effects" of marital status and schooling, respectively, on the gender wage gap. For example, the "never-married" row is the gender wage gap implied by the regression coefficients for the hypothetical case that all FTFY women were never married, holding constant all other demographic characteristics and the employment rate. Never-married status is associated with a gender gap that is 0.225 smaller (0.225 = –0.179 + 0.404) than average in the 1970s. Conditional on never-married status, the OLS gender gap closed much less than did the average OLS gender gap. The selection-corrected gap closed about the same for the various marital groups, ranging from –0.066 to +0.019. Overall, Table II suggests that selection bias increased the least for the never-married group—the group for which the OLS gender wage gap closed the least. Similarly, among the schooling groups, selection bias increased the least for those with advanced degrees, which is the group with the highest employment rate and a relatively small change in the OLS gap.

Figure III displays uncorrected estimates, two-step estimates, and a 95% confidence interval for the uncorrected, for all of the five-year intervals between 1970 and 1999, using variable weights. According to the two-step estimates (marked with squares), the gender wage gap has been essentially constant within groups during these years, even though the uncorrected gaps (solid line without markers) have closed within groups. The difference between the two series suggests that part of the measured gender-gap closure comes from changes in the composition of the female workforce (relative to the male workforce) within demographic group. The 95% confidence intervals for the uncorrected gap do not overlap two-step estimates in the 1970s or in the 1990s.

III.C. Some Basic Patterns in the Wage Data Suggesting That the Selection Rule Has Changed

The gap between the marginal and average worker becomes less relevant as the employment rate L approaches 1, which is why $\lambda' < 0$ and $\lambda(1) = 0$. If much of the measured wage change for women is due to changing selection bias, then measured wage changes should be less for demographic groups with high and stable employment rates as compared to groups with low employment rates, because selection bias is close to zero (and therefore cannot change significantly) when the employment rate is high.

Recall from the inequality (5) that, even without the normality assumption, the GHR model features two kinds of changes in the composition of the workforce. One is a uniform change in the employment threshold for all workers (the right-hand side of (5)) that, assuming selection is "positive," moves the employment rate and measured average log wage in opposite directions because the marginal worker has lower wages than the average worker.[17] The second reason the workforce composition can change is through the selection rule. An increase over time in the return to skill, for example, can pull more skilled women into the workforce and push less skilled women out. In this case, we do not expect a negative relationship between measured wages and employment rates. Instead, we expect measured wages to grow for women—even though their employment rates may be increasing. This is especially true for those groups of women with low initial employment rates, because $\lambda(1)$ is always 0, even though the selection rule is shifting over time.

Figure IV explores the relationship between the *level* of the employment rate and the change in the gender gap, across demographic groups defined in the same way as in our regression analysis.[18] Each marker in the figure represents a demographic group, with the never-married (ever-married) groups indicated by squares (circles). The labels distinguish the groups by education and years of potential experience (midpoint of the year range 5–14, etc.). Figure IV shows a negative relationship. The female groups with the highest initial employment rates (experienced ever-married women with advanced degrees and various groups of never-married women) had hardly any relative wage growth. Less experienced women with advanced degrees had employment rates in the middle of the range, and had relative wage growth of about 0.11 log points. Ever-married women without advanced degrees had the lowest initial employment rates, and the highest relative wage growth. Together with the

17. Olivetti's and Petrongolo's (2007) recent paper features this comparative static and finds that countries with higher female employment rates have lower female relative wages.

18. To make Figure IV legible, we combine some of the categories from the regression, namely, one less-than-high-school-degree category (rather than distinguishing high school dropout from no high school attendance), two marital categories (never-married and ever-married) and three potential experience categories (5–14 years, 15–24 years, and 25–34 years). In addition, we drop demographic groups for which our CPS wage sample includes fewer than forty female observations per year during the 1975–1979 period. Potential experience is measured as Age – Schooling – 7.

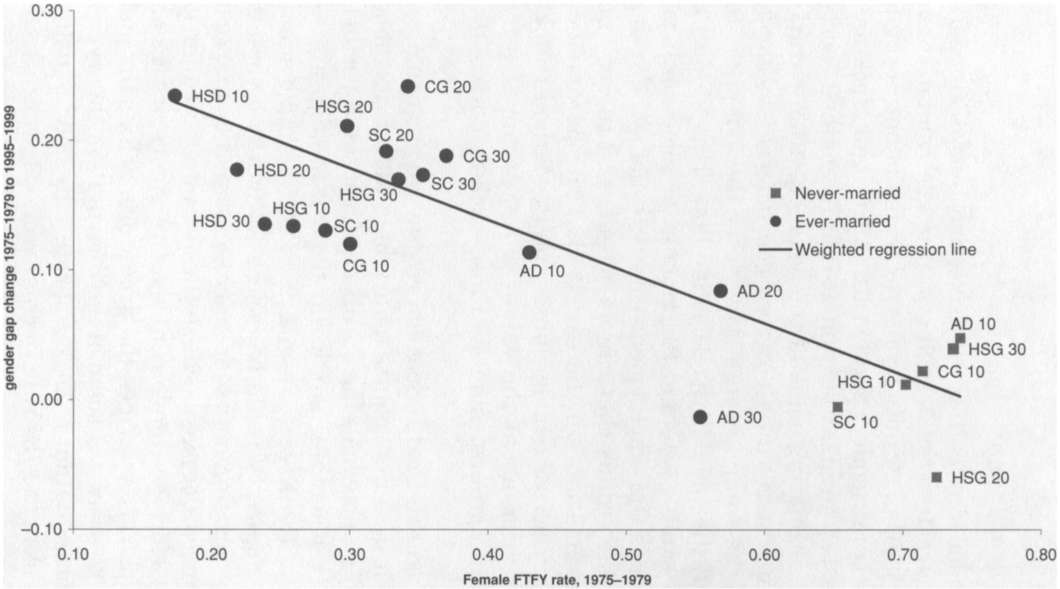

FIGURE IV

Measured Wage Growth Declines with Labor Supply

The scatter diagram shows the gender-gap change 1975–1979 to 1995–1999 against the FTFY employment rate 1975–1979 for the 21 demographic groups with at least 40 observations of female FTFY workers per year in the 1970s. The demographic groups are the cross-product of marital status (never-married vs. ever-married), schooling (high school dropout, HSD; high school grad, HSG; some college, SC; college grad, CG; and advanced degree, AD), and potential experience (10 denotes 5–14 years, 20 denotes 15–24 years, and 30 denotes 25–34 years). The calculations use our CPS sample of white persons aged 25–54, trimming outliers and adjusting topcodes as described in Appendix I.

Heckman two-step estimates, these results suggest to us that selection bias is potentially important, and should be measured with methods that are consistent with a time-varying selection rule.

III.D. Identification at Infinity

The GHR model implies that the selection bias disappears for groups with characteristics **X** such that practically all of them work, even if the joint distribution of error terms is not normal. On the basis of this principle, Chamberlin (1986) and Heckman (1990) have suggested an identification-at-infinity method (hereafter, Method II): estimating some of the wage equation parameters using a sample *selected on observed characteristics* **X** such that nearly all of the sample works. The closer the sample is to 100% employment, the smaller the selection bias. This implies a trade-off between sample size and the amount of selection bias, because the higher is the employment-rate threshold, the fewer are the demographic groups that can surpass it and thereby be included in the estimation. To the extent that the parameters of interest $\{\gamma_t\}$ vary across demographic groups, Method II does not provide information about the values of the parameters for several demographic groups, namely, those excluded from the estimation.

To its advantage, Method II does not require an exclusion restriction; namely, it does not require a variable that affects labor supply without affecting wages (i.e., is absent from the **X** vector). Instead, Method II designates particular demographic groups—particular values of the **X** vector—with high FTFY employment rates. In a sense, Figure IV could be used as a heuristic version of this method, by focusing on never-married women or married women with advanced degrees because these groups have fairly high FTFY rates. Here we implement a more rigorous version (Heckman [1990, p. 317] describes the procedure; Schafgans and Zinde–Walsh [2002] work out some of the distribution theory) by estimating a probit equation $P(X\delta_g)$ separately by gender using CPS observations from the 1975–1979 cross sections, where X are variables from the wage equation and $\boldsymbol{\delta}_g$ is a vector of coefficients for gender g. The dependent variable is working FTFY. This probit equation serves *only* the purpose of selecting demographic groups to be included in a wage equation: we select only CPS men and women who are employed and have demographic characteristics such that predicted probability exceeds α, for α close to 1.[19] A selection-bias-corrected gender wage gap time series is then calculated as the conditional gender gap for these observations only. Specifically, the regression equation for a cross section t is

$$(15) \qquad (w_{it} - X_{it}\beta_t) = g_i\gamma_t + \varepsilon_{it}^w, \quad \{i \mid P(X_{it}\delta_g) \geq \alpha, L_{it} = 1\}.$$

By construction, the restriction $L_{it} = 1$ by itself has little impact on the sample used to calculate the conditional gender gap, because most of the persons in the demographic groups satisfying the threshold are also employed themselves. To calculate the dependent variable for the regression (15), estimates of the coefficients on the other demographic variables are needed (Heckman 1990; Schafgans and Andrews 1998). Since our paper has not given much attention to the selection bias (if any) of the other coefficients, the results displayed here take the coefficients from the OLS regression (which does not include the inverse Mills ratio or any other result from the probit equation) on the full sample of persons working FTFY.[20]

Figure V's square-marked series displays estimates of $\{\gamma_t\}$ using (as in Figure III), cross sections formed by pooling five adjacent waves of the CPS, but keeping only observations with $P(\mathbf{X}\delta_g) \geq 0.80$ and $L_{it} = 1$. It shows a gender gap of about –0.20 log points that does not trend over time. Recall that the uncorrected gender gap closes more than 0.15 log points (shown in the figure as a solid series with no markers).

If the assumptions of Method II are correct, the amount of selection bias falls as the employment rate threshold (used to exclude demographic groups from the wage regression) rises. Figure V illustrates this point with the circle-marked, triangle-marked, and unmarked series, which use 0.70, 0.60, and 0 employment rate thresholds, respectively. Of the series shown, the square-marked and circle-marked series (80% and 70%, respectively) trend the least, the triangle-marked (60%) trends upward second most, and the unmarked series trends upward the most. One interpretation of these results is that the upward trend is

19. For $\alpha = 0.80$, 0.5%, 0.7%, and 1.2% of the white female FTFY CPS observations aged 25–54 satisfy these criteria in the 1970s, 1980s, and 1990s, respectively, for a total of about 300 female observations per five-year cross section. Because of this relatively large number of observations, we are able to consider thresholds closer to 1 than does Schafgans' (1998) wage gap study.

20. We have also tried taking the other coefficients from our two-step estimates or, most ambitiously, to also estimate them with the infinity method. These alternative approaches give very similar estimates of selection-bias changes.

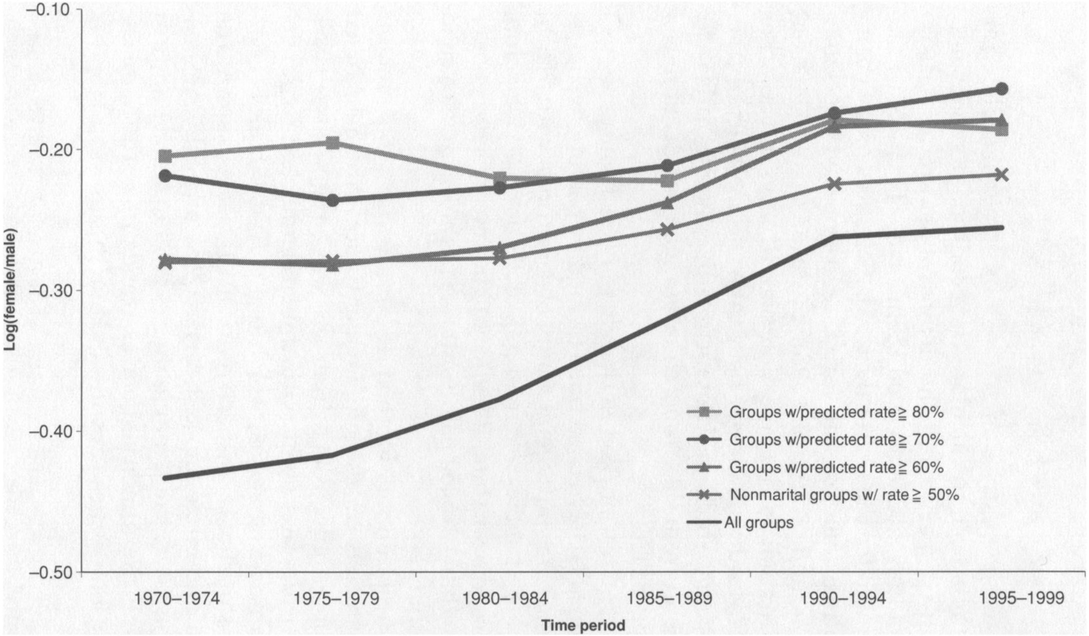

FIGURE V

Gender Wage Gaps Among Strongly Attached Groups, Various Thresholds

The figure graphs five times series of women's log wages relative to men's, net of demographic characteristics. The series differ according to the demographic groups (defined according to gender, schooling, marital status, and potential experience, except for the x-marked series that does not use marital status) included in the estimation. The unmarked series includes all demographic groups. For the other series, demographic groups are selected based on their FTFY employment rate for the years 1975–1979. The calculations use our CPS sample of white persons aged 25–54, trimming outliers and adjusting topcodes as described in Appendix I.

the result of changeable selection bias, which can be mitigated by using Method II with a high threshold.

Because marital status and the age at first marriage have changed over time and may themselves react to the wage structure, it is possible that the presence of marital-status regressors biases our results. To explore this possibility, we prepare Method II results without any marital-status regressors, although doing so requires reducing the threshold.[21] Figure V's x-marked series is based only on the schooling, region, and potential experience regressors and a 50% FTFY rate threshold. With a threshold so far below 100%, we expect that part of the x-marked series' increase is due to changing selection bias, but nonetheless it increases only 0.062 from 1970–1974 to 1995–1999.[22] Thus, our Method II conclusion that the selection bias has been a major factor in the narrowing of the gender wage gap does not depend on the use of marital status as a regressor.

IV. Possible Implications of the Wage Evidence for Measuring the Effects of Labor Demand Factors

The economic literature has proposed two repeated cross-sectional methods for measuring the change in selection *bias* over time in an environment in which the selection *rule* changes over time: control function methods and Method II.[23] On the basis of results from the two methods, we can help to answer some broader questions about the factors influencing women's relative wages. First, we acknowledge and interpret each method's limitations. Second, we explain how the average man may not be the appropriate benchmark for the purpose of measuring gender-specific differences in labor demand.

The Heckman two-step and identification-at-infinity methods are based on different assumptions and data requirements. When implementing the Heckman two-step method, it is desirable to

21. Recall from Figure IV that the highest FTFY demographic groups are never-married.

22. The x-marked series may be best compared with the circle-marked series (70% threshold), because both include 3–6% of the female FTFY workforce and both have an average female log wage of 2.7 in 1975–1979. The circle-marked series also increases 0.062 from 1970–1974 to 1995–1999.

23. A panel data method for measuring the change in composition over time has been proposed in the business-cycle literature (Bils 1985; Solon, Barsky, and Parker 1994). This method is also consistent with a changing selection rule; future work may adapt it to the question of long-term changes in the gender wage gap (see Wellington [1993] for such a study of nine-year changes).

have an "instrumental" variable that affects selection but not wages.[24] Method II does not use an instrumental variable (and permits selection corrections for both women and men), but assumes that the parameters to be estimated (in our case, gender differences γ in labor demand) do not vary across demographic groups. It requires large samples. Given that we have applied both methods to repeated cross-sectional data, one common weakness for our purposes is that they do not include measures of historical work experience that have been shown to be important for explaining the level of the gender gap (Mincer and Polachek 1974; O'Neill and Polachek 1993).[25] Method II results are probably less vulnerable to this criticism because they are based on comparisons of men and women with similar degrees of labor force attachment, although the comparisons are limited to those with very strong attachment.

Figure VI compares results from the two methods, displaying indices for women's relative wages using 1975–1979 as a base period. For example, a value of 110 means that, according to one of the methods, the female-male wage ratio was 10% greater than it was in 1975–1979. The unmarked series uses no selection correction, showing significant relative gains for women. The other series are based on selection corrections—one series for each of the methods—and show very little relative gains for women. Although each of the methods is based on its own debatable assumption—an exclusion restriction or assumption of uniform gender gaps across groups—the methods deliver similar estimates of selection-bias changes since the 1970s. This finding suggests, but does not prove, that our results about changes over time are not driven by any particular one of the methods' selection-correction assumptions.

24. Following the previous literature, we have taken (for women) the instrumental variable to the number of young children in the household. See Mulligan and Rubinstein (2005) for a discussion of how possible endogeneity of children might affect estimates of selection-bias change. One disadvantage of instrumental variable approaches is that a good instrumental variable for men's labor supply may be even harder to find.

For an exploration of the robustness of selection-bias change estimates to alternative instrumental variable specifications, see Mulligan and Rubinstein (2004). In some cases, the Heckman model can be estimated without an instrumental variable by exploiting the nonlinearity of the inverse Mills function.

25. Blau and Kahn (2006) show that a narrowing of the gender gap in historical work experience explains some of the gender-gap closure in the 1980s and 1990s, but leaves a large majority of it unexplained. O'Neill and Polachek (1993) have a similar finding for the years 1976–1987. Regardless of how much gender wage-gap closure might be attributed to narrowing of the gender experience gap, our results still support the broader conclusion that much of the gender wage-gap closure should be attributed to supply factors.

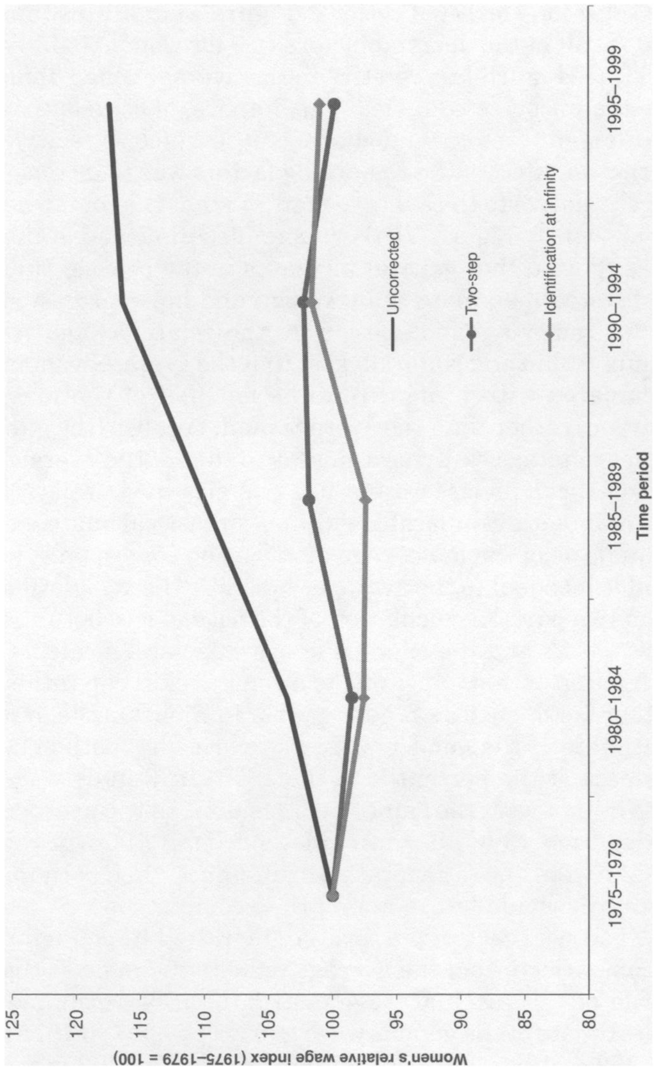

FIGURE VI

Gender Relative Wage Indices with and without Selection Corrections

The figure graphs three time series of indices of women's wages as a ratio to men's (1975–1979 = 100), net of demographic characteristics. The series differ according to the method for correcting selection bias. The calculations use our CPS sample of white persons aged 25–54, trimming outliers and adjusting topcodes as described in Appendix I.

The gender wage gap, and therefore its changes, reflects a combination of factors including different amounts and types of skills possessed by men and women (supply factors), different market valuations of those skills, and discrimination (demand factors). The selection-corrected series in Figure VI indicates that the combined result of the demand factors was for women to have wages that changed much less relative to the average man's than suggested by the uncorrected series. This implies, of course, that if some of the demand factors tended to decrease women's relative wages, then the net effect of the remaining factors was to increase women's wages relative to the average man's so that the combined result is as shown in Figure VI. For example, Blau and Kahn (1997) have suggested that general increases in the price of skill should have favored men more than women and have suggested a procedure for removing this effect from the relative wage series.[26] Put simply, Blau and Kahn suggest that the average woman should be compared with a man from the left half of the male wage distribution, rather than the average man, because (they argue) the former man has skill more similar to that of the average woman. To construct a precise estimate of women's wages relative to those of a man with comparable skill, one must calculate exactly which quantile of the male wage distribution corresponds to a general skill level equal to the average woman's. The calculation can be done in two parts: a calculation of the *level* of gap between the average woman's and the average man's wage and a calculation of what fraction of that gap is due to a general skill gap rather than some other factor such as gender discrimination. On the first point, our estimate of the gender wage gap varies by method so that our estimate of the percentile of the 1975–1979 male wage residual distribution with the same wage residual as the average woman's varies from 13 (OLS) to 27 (Method II). If 100% of the gender wage gap were due to general skill difference, then perhaps the 20th percentile would be the appropriate comparison.

Figure VII adjusts each of the series in Figure VI by changing the basis of comparison from the average male to the male at the 20th percentile of the residual wage distribution. For example, the two-step estimate of the gender wage gap was –0.337 in 1975–1979 and –0.339 in 1995–1999, which makes a 1995–1999 index

26. Juhn, Murphy, and Pierce (1991) used the same procedure to show that black men gained rank in the white male wage distribution during the 1980s, even though black men's relative wages were fairly constant.

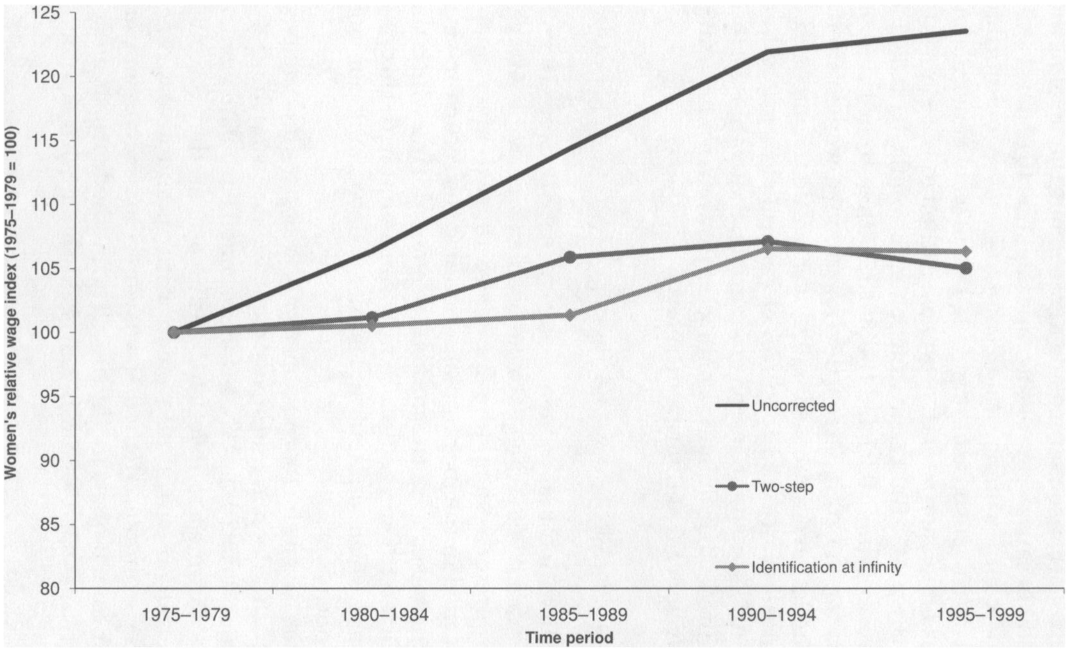

FIGURE VII

Indices of Women's Average Wages Relative to the 20th Percentile Man

The figure graphs three indices (1975–1979 = 100) of women's average wages as a ratio to the wage at the 20th percentile of the male wage distribution, net of demographic characteristics. The series differs according to the method for correcting selection bias. The calculations use our CPS sample of white persons aged 25–54, trimming outliers and adjusting topcodes as described in Appendix I.

of 99.8 in Figure VI for the average woman relative to the average man. However, the 20th-percentile man in the CPS wage residual distribution lost 0.051 log points of wages relative to the average man, and so Figure VII's 1995–1999 index is 105.0, or 99.8 divided by 0.95.[27] Because methods disagree about the level of the gender gap and because some of the gap is not due to general skill differences, percentiles higher than 20 might be more appropriate comparisons for the purpose of removing the effect of general skill price changes on women's relative wages. Although not shown in Figure VII, the 1995–1999 adjustment factor would be 0.96 for the 27th percentile and 0.97 for the 35th, rather than the 0.95 we calculated for the 20th percentile. In summary, women's relative wages would have increased 0–0.05 log points if composition of the female workforce and the level of general skill prices had been constant. This slight increase suggests that other factors may have shifted labor demand somewhat in women's favor and were offset by general skill price increases.

V. Investment Interpretations of Selection Corrections

Both observation and economic theory (e.g., Becker [1985b]; Goldin and Katz [2002]) suggest that women's growing attachment to the workforce should have induced them to invest more in skills rewarded in the marketplace, because the return-to-skill investment increases with the number of hours during which the skills are utilized.[28] Our results are fully consistent with this possibility, because "selection" and "investment" have many common economic implications.

To discuss this issue more precisely, first notice that the possibilities of both selection and investment create two potentially interesting counterfactuals. What would women earn if all women worked, but continued to invest (or not invest) as they do now? What would women earn if all women worked, and invested accordingly? For example, Method II answers the second counterfactual better than it answers the first. The method focuses on demographic groups (such as never-married women with advanced

27. The adjustment factor varies by year according to the gap between the 20th percentile and the average. Most of the widening of that gap occurred prior to 1985 (see also Autor, Katz, and Kearney [2005]).

28. Some of their investment would go unmeasured in the CPS (e.g., the subjects that women study in school—see Grogger and Eide [1995]—or their commitment to advancing their careers), and thereby result in the closure of measured conditional gender wage gaps.

degrees) in which almost all women work. If these women expected to be working, then we expect them to have invested accordingly. In this sense, Method II "corrects" for both selection and (lack of) investment by calculating wages for groups of women within which labor force selection is less important (essentially all of them work) and for whom investment rates are more like men's.[29]

Because human capital investment is complementary to labor force attachment, investment may respond to a changing selection rule, and the selection rule might change in response to a general increase in the returns to investment to the extent that returns to investment vary across women. Suppose, for example, that it has always been true that, for a given attachment to the labor force, high-IQ women have a greater return to human capital investment than low-IQ women. In the 1970s, low-IQ women intended to work and high-IQ women did not. As a result, investment was zero for high-IQ women and positive, but (because of their low returns) small, for low-IQ women. Both because of the direct effect of IQ on wages and because of the indirect effect via investment, the gender gap among workers was quite large in the 1970s. By the 1990s, high-IQ women intended to work as much as the low-IQ women did (and did in the 1970s), but they invested more because of their high returns. The gender gap shrank between the 1970s and 1990s both because high-IQ women entered the workforce and because high-investment women entered the workforce. Method II (as we have implemented it by following the same demographic groups over time) would not measure much of this source of women's measured wage growth.

Complementarity between human capital and labor force attachment also implies that a general increase in the returns to investment might differentially pull high-IQ women into the labor force. For our purposes, the lesson is that women's wages may have grown because their labor market behavior increasingly resembles male behavior, and that selection-correction methods consistent with time-varying control functions can help measure wages for a counterfactual world in which women work and invest as men do. We leave it to future research to carefully and quantitatively contrast investment and unobserved selection. The wage

29. See Appendix I of Mulligan and Rubinstein (2005) for a formal model showing how control-function methods (such as the Heckman two-step estimator) might also simultaneously "correct" for selection and investment.

evidence presented in this paper suggests that women's measured relative wages have grown in large part because they behave differently than they used to in terms of labor supply, labor force attachment, human capital investment, or some combination of these factors, even if the evidence does not yet show the relative quantitative importance of various behavioral changes.

VI. Nonwage Evidence of the Changing Composition of the Female Workforce

The basic economic logic of the GHR model applies just as well to observed human capital proxies as to unobserved human capital. For our purposes, one important conclusion from the GHR model is that selection could be negative: the female workforce could have less skill than the female population. The second conclusion is that an increase in the return to human capital may affect the supply of skilled women proportionally more than it does the supply of unskilled women.[30] If so, then the skill composition of the female workforce will be greater than it used to be even if skill were a fixed factor at the individual level. Here we cite two examples of possible negative selection, which over time became less negative (or became positive).

Figure VIII displays the FTFY employment rates of married women as a function of their husband's position in the wage distribution. The first step in making the calculation is to estimate a probit equation in our CPS sample of currently married women whom we can match with husbands aged 25–54. The regressors are a quartic in the wife's potential experience, the wife's schooling dummies, the wife's schooling interacted with the potential experience quartic, region, and dummies indicating the husband's quartile in the hourly wage distribution for husbands his age in the year in which we measure his hourly wage. For the average values of the wives' regressors, we calculated a fitted value from the probit for each husband's quartile. To determine whether a quartile's employment rate increased in a greater or lesser proportion than the rate for the general population, this predicted employment rate is divided by the predicted employment rate for the

30. According to the GHR model, an increased return *may* increase the skill intensity of the labor force. As a counterexample, consider that an increased return to advanced degrees might not increase the fraction of working women with advanced degrees because the employment rate of women with advanced degrees was very high in the first place.

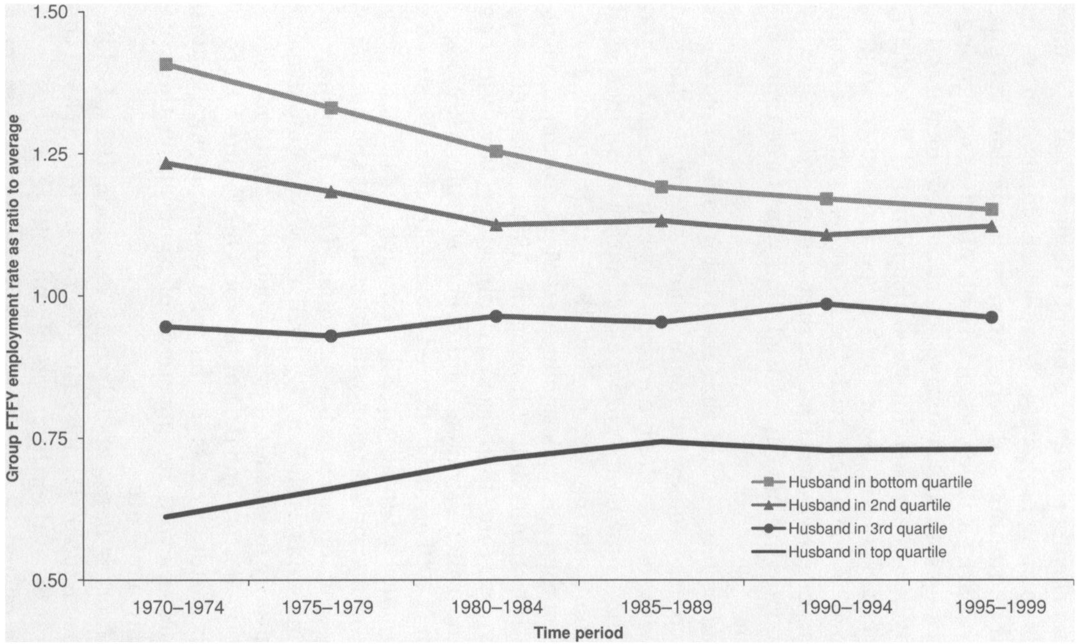

FIGURE VIII
Relative Employment Rates by Husband's Wage Quartile

The figure graphs four time series of a group's FTFY employment rate as a ratio of the FTFY employment rate of all currently married women. The groups differ according to the husband's position in the distribution of wages of husbands aged 25–54. The calculations use our CPS sample of white persons aged 25–54, trimming outliers and adjusting topcodes as described in Appendix I, and limited to persons currently married.

same average regressors from a probit equation that does not include the husband-quartile dummies. We see that women with husbands earning low wages have higher employment rates than the average married woman in each year (see the square-marked and triangle-marked series at the top of the figure) and women with husbands earning high wages have lower employment rates (circle-marked and unmarked series toward the bottom). If husbands and wives are positively assortatively matched on earnings potential, this suggests that negative selection of women into the labor force may be a real possibility (see also Blundell et al. [2007, p. 331]). The fact that the lower lines slope up, and the higher lines slope down, is consistent with the hypothesis that selection has become less negative over time, so that women with relatively high earnings potential have increased their workforce representation over time.

Other studies have found that married-female employment rates have increased somewhat more among the more educated and among women with high-earnings husbands. Aguiar and Hurst (2007) show how hours per week of market work increased significantly more during the period 1965–2003 among more-educated women than among less-educated women. Juhn and Kim (1999, pp. 30–31) explain that "[d]uring the 1980s . . . increases in female employment rates have occurred almost exclusively among high school and college women." Juhn and Murphy (1997) and Juhn and Kim (1999) stratify married women by their husband's position in the married-male wage distribution. In 1970, the employment rate of married women with husbands at the bottom of the male wage distribution was 0.44, compared to 0.31 for married women with husbands at the top. By 1990, the wives' employment rate was essentially independent of the husband's position, for example, 0.60 at the bottom and 0.61 at the top. Our Figure VIII complements their results by calculating relative employment rates, which are more relevant for questions of composition than are absolute employment rates.

Work with other data sets (Grogger and Eide 1995; Goldin, Katz, and Kuziemko 2006) shows that women have increased the market orientation of their high school and college studies (with more emphasis on math, business, etc.) relative to men. Assuming that the reorientation of female schooling has been rewarded in the labor market and has occurred in greater proportion among women who ultimately work FTFY than among women who do not, these changes would be detected in our analyses as changes

TABLE III
EFFECT OF IQ ON A WOMAN'S LIKELIHOOD TO BE A FULL-TIME FULL-YEAR WORKER

		Probit	
Variables	OLS (1)	coeff. (2)	dF/dX (3)
IQ above 100	−0.006	−0.023	−0.009
(1968–1979)	(0.019)	(0.063)	(0.024)
IQ above 100* 1980s	0.040	0.119	0.046
	(0.021)	(0.065)	(0.026)
IQ above 100* 1990s	0.062	0.168	0.065
	(0.026)*	(0.076)*	(0.030)*
Observations	21,308	21,308	21,308
Number of individuals	2,135	2,135	2,135

Notes. The table reports some of the coefficients from a linear regression (column (1)) and a probit equation (columns (2) and (3)), each with binary dependent variable equal to 1 for those reporting working 35 hours per week and at least 50 weeks of the year. IQ is a composite of various test scores measured by the Census Bureau's school survey (see Appendix I). In addition to the three regressors shown in the table (by row), the equations include demographic variables interacted with decade. Column (3) reports the marginal effects for the column (2) probit, evaluated at the sample mean.

Person-years are the unit of observation. The calculations use our NLSW sample of white women aged 25–54, for the calendar years 1969–2000 (all of these women were 14–24 years of age in 1968).

Robust standard errors are in parentheses.

* significant at 5% for two-sided hypothesis.

in selection bias on unobservables, because the degree of market orientation of a person's schooling is unobserved in the CPS.

Table III reports results from the NLSW on yet another characteristic unobserved in a CPS study: test scores. The microdata we use are from the young-women sample, 14 to 24 years of age in 1968, that covers the years 1968 to 2003. Our sample includes white women of ages 25 to 54 during the working year, which means that our subsample covers the working years 1969–2000. We exclude individuals enrolled in school in the given year and observations with missing report on hours or weeks worked. We further exclude observations with missing test score data or labor market outcomes (see details in Appendix I). The test score data were collected by the Census Bureau via a separate school survey, transformed to a composite score, and referred to as the "IQ Score" (see Appendix I).

One approach to measuring the change in selection on cognitive skills, unobserved by the econometrician in larger representative data sets, is to estimate the correlation between IQ test score and work status in the 1970s, the 1980s, and the 1990s. We group respondents by education, marital status, and region of residency as we do with the CPS. As in our log hourly wage specifications, we allow labor supply to vary also by potential experience (quartic) and differently over education groups. Our labor-supply regressions pool all years together, allowing the regression functions to vary by decades. Table III presents the estimates of the effect of IQ on the likelihood of working FTFY for the 1970s, 1980s, and 1990s.[31] The various columns report results from two models of the probability that a woman works FTFY (one linear, the other probit). The probit coefficients are reported in two formats: as coefficients on the probit index (middle column) and as marginal probability effects (last column). Column (1)'s linear regression and column (2)'s probit regression both show that high-IQ women were somewhat less likely in the 1970s to work FTFY, although the effect is statistically insignificant. The relationship between IQ and FTFY status is economically and statistically significantly different in the 1990s than in the 1970s. For example, the probit model suggests that FTFY employment rates increased 6.5 percentage points more over time among high-IQ women than among low-IQ women.[32] The OLS column shows a similar pattern.

VII. Conclusions

After years of a fairly constant gender wage gap, women's measured hourly wages grew almost 0.20 log points relative to men's from the late 1970s through the late 1990s. Although previous studies have found it coincidental—even paradoxical—that wages have become more equal between genders at the same time that they have become so much less equal within gender, we suggest that growing inequality within gender, through its effect on women's selection into the labor force, their labor force attachment, and their human capital investment, is a major reason why the wages of the female workforce have grown relative to men's.

We use three different empirical approaches to attempt to measure the existence and importance of these effects: additional measurements of female workforce characteristics, Heckman's two-step estimator, and identification at infinity. Each of these

31. Our specification has IQ measured as a dichotomous variable (greater or less than 100) because of the likely significant measurement error (i.e., our specification is the reduced form of a Wald estimator; see Greene (2000) for further discussion).

32. The coefficient on the interaction term is 0.065, and so it adjusts for the fact that time changes in FTFY rates are expected to be different for the two groups merely because they began at different FTFY rates and the labor supply function is nonlinear (Ai and Norton 2003).

approaches is consistent with, but does not assume, the possibility that the composition of women's labor supply has fundamentally changed over the years. Relative to those of non-employed women, the IQ and husband's wages of employed women have risen, suggesting that female workforce selection has shifted from negative to positive, or at least has become less negative, over time. As shown in Figures VI and VII, the two other methods permit numerical adjustments for changes in selection (and some changes in investment) among women, as well as for changes in schooling and experience for the two genders. The results suggest that the majority of measured women's relative wage growth would not have occurred if it weren't for the change in the composition of the female workforce. If our adjustments are correct, they imply that women's wages grew between 0.05 and 0.06 log points relative to the 20th percentile man's, and even less relative to the average man's, from the late 1970s to late 1990s.[33]

The GHR model suggests an explanation of why wages might become more equal between genders at the same time that they became less equal within gender, namely, that inequality within gender ought to affect the selection rule and the distribution of human capital investment among women and thereby the composition of the female workforce. This changing composition of the female workforce is at least partly responsible for the closing of the gender gap. More technically, the model says that inequality affects the coefficient on the inverse Mills ratio in a log female wage regression, and that the selection bias is the combination of this coefficient and the inverse Mills ratio itself (which has declined over time due to rising female employment rates—see our equation (12)). Figure IX graphs the inverse Mills ratio coefficient—estimated once for each annual cross section using the same Heckman two-step specification as in Table I—as a circle-marked series. Because the coefficient on the inverse Mills ratio has increased over time and the increase is an estimate of what would have happened to selection bias if the inverse Mills ratio had been held constant, the circle-marked series is inconsistent with the hypothesis that the selection rule for unobserved characteristics has been constant over time.

The Heckman two-step estimator conditions on observable characteristics. Identification at infinity is also designed to

33. The final values of the women's relative wage indices (1975–1979 = 100) are 99.2 and 100.9 in Figure VI (average male benchmark) and 105.0 and 106.3 in Figure VII (20th percentile male benchmark).

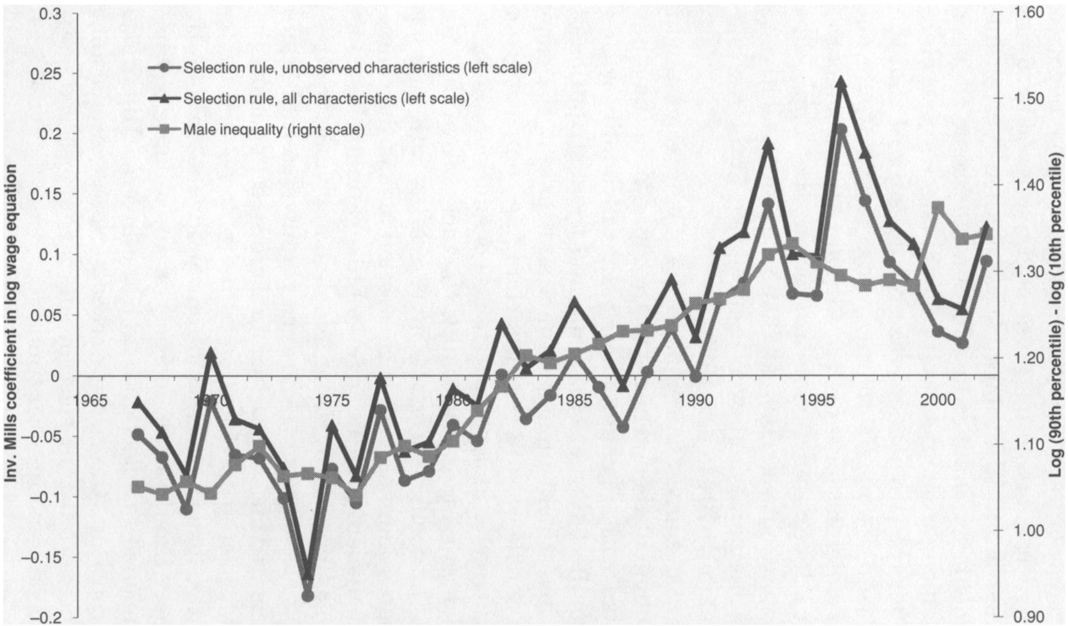

FIGURE IX
Empirical Measures of the Selection Rule and Wage Inequality

The figure graphs time series of (a) the selection-rule coefficient for unobserved characteristics (left scale), (b) the selection-rule coefficient for all characteristics (left scale), and (c) the log of the ratio of the wage of a man at the 90th percentile of the male wage distribution to that of a man at the 10th percentile (right scale). The former selection-rule coefficient is the coefficient on the inverse Mills ratio in a cross-sectional log female wage equation. The latter selection-rule coefficient is the gap between predicted wages for FTFY employed and all women, divided by the average inverse Mills ratio for FTFY women.

correct for selection on unobserved characteristics. Nevertheless, the economic logic of the GHR model applies equally well to observed as to unobserved characteristics. Thus, it is important to know whether and how the selection rule has changed for all characteristics combined and not just the unobserved ones. The triangle-marked series in Figure IX graphs the overall analogue to the conditional selection rule displayed as the circle-marked series. Specifically, each value of the triangle-marked series is the gap between working women's log wages and all women's log wages predicted by the Heckman two-step model for that cross section, divided by the average inverse Mills ratio for working women.[34] The triangle-marked series is always slightly above the circle-marked series, which means that overall selection bias is always slightly more positive (or slightly less negative) than selection bias for unobserved characteristics. More important, the overall selection rule follows essentially the same time trend as the selection rule of the unobservables, if not changing a bit more over time.

Figure IX also reproduces the within-gender inequality measure from Figure I, now as a square-marked series. Like the selection rule indicators (circle- and triangle-marked series), wage inequality increased less during the 1970s, more during the 1980s, and has had little (if any) upward trend since the early 1990s. As compared to the time series for male wage inequality, the timing for the selection-rule changes shown in Figure IX is suggestive of a possible causal link, but much more work needs to be done to link inequality within gender to the composition of the female workforce. If the link is important, our results suggest that the measured gender wage gap will remain as small as it currently is only as long as inequality remains at its currently high levels.

Although we attribute the majority of women's relative wage growth to the changing composition of the female workforce, there are a couple of reasons to suspect that changing discrimination and increases in the relative price of skills possessed relatively intensely by women may have pushed in the direction of raising women's relative wages. First, a finding that women's wages have grown like the average man's is consistent with women's gaining rank in the male wage distribution because, if their rank had been constant, then their wages might have fallen relative to

34. For the purpose of calculating the average log wage for working women, log wages for working women with missing wages are imputed as regression-fitted values.

the average man's. Blau and Kahn (1997) explain that working women in the 1970s were typically earning less than the median man, and that persons (even male persons) earning less than the median man had their wages fall over time relative to the median man's.[35] Second, other forces may have pushed toward relative reductions in women's wages, thereby counteracting any gains they might have enjoyed from changing discrimination or increases in the relative price of women's skills. For example, male and female labor may be imperfect substitutes in production, so that women's relative wages were (other things constant) reduced by the growing size of the female workforce. In summary, even if we are correct in our finding that women's relative wages would have risen modestly absent changes in the composition of the female workforce, it is possible that other factors are needed to explain why women's relative wages rose modestly rather than fell.

Appendix I: Formation of, and Summary Statistics for, the Microdata Samples

Our CPS observations are white non-Hispanic adults between the ages of 25 and 54, excluding persons living in group quarters or with missing data on relevant demographics. Our estimates of wage equations further limit the sample to persons working FTFY, excluding the self-employed; persons in the military, agricultural, or private household sectors; persons with inconsistent reports on earnings and employment status; and persons with allocated earnings. We trim wage outliers when calculating mean wages, and keep them when calculating quantiles (as in Figure I). The purpose of this appendix is to give more detail about these samples and their variable definitions.

About 15 percent of CPS respondents refuse to answer the earnings questions in the March CPS supplement. Therefore, the Census Bureau uses a matching-on-observables method—a "hot deck" procedure—to impute or allocate earnings for the nonrespondents, based on the response for a sample person with similar demographic characteristics. Our wage samples exclude these respondents. When calculating means, or mean regressions, we trim outliers according to the measure of hourly wages: excluding observations with measured hourly wage below $2.80 per hour

35. We calculated this effect between 1975–1979 and 1995–1999 to be 0–0.05 log points.

(year-2000 dollars, as deflated by the Consumer Price Index) and below the 1st-percentile value of the FTFY male hourly wage distribution (this distribution includes wages for nonwhite never-married men aged 18–65) or above the 98th-percentile value of the same distribution.[36] The trimming almost completely eliminates persons who had their annual earnings topcoded by the Census Bureau.[37] The first six columns of Table A.1 report summary statistics for the female members of the main sample and by gender for members of the wage samples. The next two columns of Table A.1 display average characteristics for female and male members of our identification-at-infinity sample: FTFY workers who are members of a demographic group with FTFY predicted employment rate at least 80% for the years 1975–1979 (below, we explain how the employment rates were predicted).

CPS measures of education and annual hours are not standardized over time. Beginning with the March 1992 CPS, education was no longer classified according to years of school, but according to the highest degree received classified into categories. Therefore, we classify the adult population into six educational categories: (i) persons with 0–8 years of schooling completed; (ii) high school dropouts; (iii) high school graduates, 12 years of schooling (including GEDs); (iv) some college, which prior to the 1992 survey refers to 13–15 years of completed schooling and since then "some college" but no degree, or associate's degree; (v) college graduate, which prior to the 1992 survey refers to 16 years of schooling completed and since then bachelor's degree; and (vi) advanced degree, which prior to the 1992 survey means at least 18 years of schooling and since then a master's, professional, or Ph.D. degree. Potential work experience is constructed as the max between zero and age (in year of survey) minus years of schooling completed minus 7, where years of schooling completed

36. The first and 98th percentiles vary from year to year, but are typically 3–4 year-2000 dollars per hour (first) or 45–60 dollars per hour (98th).

37. Total annual earnings were censored by the Census Bureau at $50,000 for the years 1967–1980, at $75,000 for the years 1981–1983, and at $99,999 for the years 1984–1986. Topcoding of annual earnings for the years 1987 and following is more complicated because earnings from an individual's main job and all other jobs are censored separately, and beginning with 1995 earnings a new topcoding algorithm in the March CPS was adopted. Our trimmed sample excludes almost all of them: a few of them remain in the years circa 1979, but amount to no more than a few tenths of one percent of their cross sections. The few that remain have their earnings imputed as described by Autor, Katz, and Kearney (2005) through survey year 1995; thereafter the Census Bureau already codes the earnings for persons with censored earnings according to the average earnings for the group of CPS respondents that are censored.

TABLE A.1
CPS SAMPLE AVERAGES, WHITES AGED 25–54, NOT IN GROUP QUARTERS

	Female	Female	Female	Female	Male	Male	Female	Male
	All CPS		CPS FTFY				CPS FTFY, p80	
	1975–1979	1995–1999	1975–1979	1995–1999	1975–1979	1995–1999	1970–1999	1970–1999
Total observations	116,843	102,395	27,656	41,062	57,457	54,565	1,775	187,080
Weight	0.217	0.245	0.219	0.246	0.221	0.252	0.226	0.222
FTFY dummy	0.317	0.511	1	1	1	1	1	1
Log real hourly wage			2.487	2.607	2.924	2.861	2.971	3.016
Inverse Mills			0.997	0.727	0	0		
Widowed	0.031	0.018	0.037	0.017	0.005	0.004	0	0.001
Divorced	0.085	0.138	0.158	0.178	0.048	0.104	0.243	0.004
Separated	0.023	0.025	0.028	0.026	0.014	0.018	0	0.001
Never-married	0.063	0.118	0.128	0.151	0.082	0.156	0.757	0.001
0–8 school years	0.071	0.015	0.042	0.005	0.066	0.010	0	0
9–11 school years	0.135	0.048	0.096	0.030	0.107	0.042	0	0
High school degree	0.451	0.345	0.444	0.327	0.342	0.327	0	0.303
College degree	0.127	0.209	0.164	0.229	0.182	0.231	0.020	0.286
Advanced degree	0.032	0.089	0.055	0.106	0.085	0.104	0.980	0.136
Potential exp—15	5.073	4.527	4.112	4.139	3.564	3.742	5.840	4.341
Midwest	0.284	0.267	0.278	0.282	0.293	0.289	0.179	0.335
South	0.303	0.330	0.334	0.343	0.301	0.331	0.542	0.308
West	0.176	0.197	0.170	0.182	0.169	0.188	0.112	0.118
Kids 0–6	0.354	0.324	0.158	0.207	0.446	0.351	0.020	0.471

Note. Omitted groups are currently married, some college, 15 years' potential experience, and Northeast region.

are imputed for survey years 1992–2003 following Autor, Katz, and Kearney (2005).[38]

Prior to the 1976 survey, the March CPS surveys did not ask about usual weekly hours worked for the prior year, and reported weeks paid in broad intervals. Following Katz and Autor (1999), we impute annual weeks paid (prior year) to respondents to the 1968–1975 surveys using the 1976–1978 survey average weeks paid (prior year) within weeks paid–gender–race categories. We impute usual weekly hours worked (prior year) to respondents to the 1968–1975 surveys using the 1976–1978 survey average usual weekly hours worked (prior year) within weeks paid–gender–race–full–time categories.

Our NLSW sample derives from the National Longitudinal Survey of Young Women, which was selected to represent all women 14 to 24 years of age and living in the United States in 1968. We use the years 1968–2003, and exclude blacks and other non-white individuals. We further exclude observations with missing data on educational attainments or measures of labor supply.

The Census Bureau, via a separate school survey, collected information on NLSW participants' performances on various aptitude and intelligence tests, as well as their absenteeism and school disciplinary records. Using results from the Otis/Beta/Gamma, the California Test of Mental Maturity, and the Lorge-Thorndike Intelligence Test, as well as the PSAT, SAT, and ACT college entrance examinations, composite "IQ scores" were constructed.[39]

CPS and NLSW respondents are all classified in six educational categories: (i) 0–8 years of schooling, (ii) high school dropouts, (iii) high school graduates, (iv) some college, (v) college graduate, and (vi) advanced degree.

Appendix II: Employment and Wage Functions from Repeated Cross Sections

Table A.2 displays estimates of reduced-form female labor supply functions—one for the late 1970s (column (1)) and another

38. For survey years 1992–2003, years of schooling completed are imputed from the reported education category, gender, and race based on the sample average from prior surveys.

39. There may be psychometric problems in constructing an IQ measure from a variety of test forms. According to the NLS guide "these constructed variables were designed to keep the user who wishes to construct a unified score from having to repeat the work involved in pooling scores" (Bureau of Labor Statistics 2001, p. 58).

TABLE A.2
LABOR SUPPLY AND WAGE EQUATIONS USED FOR OLS AND TWO-STEP ESTIMATES OF THE GENDER GAP

	Effects on probability (evaluated at sample means) or regression coefficients									
	FTFY probit				log wage regression					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Gender	female	female	female	male	male	female	female	male	female	female
Years	1975–79	1995–99	1975–79	1975–79	1975–79	1975–79	1975–79	1995–99	1995–99	1995–99
Observations	116,843	102,395	116,843	112,291	57,457	27,656	27,656	54,565	41,062	41,062
Constant/predicted P	**0.297**	**0.510**	**0.307**	**0.771**	**3.027**	**2.570**	**2.665**	**2.929**	**2.663**	**2.559**
Inverse Mills							**−0.077**			**0.116**
Widowed	**0.135**	**0.024**	**0.148**	**−0.112**	−0.030	0.012	−0.009	**−0.138**	**−0.079**	**−0.072**
Divorced	**0.269**	**0.156**	**0.306**	**−0.162**	**−0.104**	**0.041**	0.003	**−0.131**	−0.010	**0.022**
Separated	**0.128**	**0.065**	**0.156**	**−0.157**	**−0.104**	−0.013	**−0.033**	**−0.078**	**−0.061**	**−0.047**
Never-married	**0.284**	**0.121**	**0.375**	**−0.181**	**−0.181**	**0.112**	**0.068**	**−0.186**	−0.003	**0.027**
Children 0–6 × constant	**−0.183**	**−0.158**								
widowed	−0.047	**0.090**								
divorced	0.008	**0.032**								
separated	0.021	0.011								
never-married	−0.062	−0.011								
Midwest	**0.029**	**0.036**	**0.028**	**0.019**	**0.008**	**−0.056**	**−0.060**	**−0.059**	**−0.099**	**−0.093**
South	**0.071**	**0.037**	**0.076**	**−0.009**	**−0.088**	**−0.112**	**−0.121**	**−0.096**	**−0.124**	**−0.117**
West	0.001	**−0.037**	0.007	**−0.058**	**0.030**	0.002	0.002	**−0.020**	**−0.027**	**−0.034**

TABLE A.2
(CONTINUED)

	Effects on probability (evaluated at sample means) or regression coefficients									
	FTFY probit				log wage regression					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0–8 school years	**−0.153**	**−0.303**	−0.188	**−0.295**	**−0.411**	**−0.381**	**−0.343**	**−0.437**	**−0.501**	**−0.566**
9–11 school years	**−0.098**	**−0.228**	−0.111	**−0.202**	**−0.272**	**−0.292**	**−0.273**	**−0.435**	**−0.421**	**−0.465**
High school degree	−0.010	**−0.023**	−0.017	**−0.061**	**−0.091**	**−0.109**	**−0.106**	**−0.174**	**−0.214**	**−0.219**
College degree	**0.021**	**0.021**	0.028	**0.047**	**0.190**	**0.198**	**0.195**	**0.277**	**0.297**	**0.302**
Advanced degree	**0.193**	**0.109**	0.219	**0.043**	**0.192**	**0.324**	**0.299**	**0.398**	**0.476**	**0.496**
(exp-15)	**−0.005**	**−0.005**	0.008	**0.005**	**0.013**	−0.002	−0.002	**0.017**	**0.009**	**0.009**
(exp-15)^2/100	**0.050**	**0.058**	0.029	**−0.064**	**−0.118**	−0.018	−0.023	**−0.091**	**−0.063**	**−0.052**
(exp-15)^3/1,000	−0.013	0.022	−0.076	**0.031**	0.025	**0.072**	**0.079**	0.023	**0.053**	**0.048**
(exp-15)^4/10,000	0.001	**−0.024**	0.027	−0.010	0.006	**−0.027**	**−0.029**	−0.003	−0.022	**−0.023**

Notes. Columns (1)–(4) are probit estimates of the propensity to work FTFY. Columns (1) and (2) are the basis for the Heckman two-step estimates in Tables I and II. Fitted values from columns (3) and (4) are used to select the samples for identification at infinity (Figure V). Columns (5)–(10) are wage function estimates. All specifications include schooling–experience interactions (not shown). Boldface type indicates statistically significant coefficients at the 95% confidence level. A bold probit constant indicates that the overall probit equation is statistically significant. The benchmark groups are currently married (living with spouse), some college, 15 years' potential experience, and Northeast region.

for the late 1990s (column (2)). These functions are the first stage of the Heckman two-step estimates displayed in Tables I and II. Columns (3) and (4) of Table A.2 display the probit equations used to select demographic groups for inclusion in the identification-at-infinity samples.

Table A.2 also displays the second-stage wage equation estimates by gender and time period. For women, second-stage estimates are shown both with and without the inverse Mills ratio as a regressor. To construct entries in Table I and the Panel A entries in Table II, we first calculate the average values of the regressors for women employed FTFY in the relevant time period (variable weights) or in the pooled years 1975–1979, 1995–1999 (fixed weights). Using these average values, we then calculate fitted values corresponding to the six wage regressions shown in Table A.2's columns (5)–(10). The OLS entries are the difference between (a) the fitted value for the female wage regression (without the inverse Mills ratio) for the relevant time period and (b) the fitted value for the male wage regression for the relevant time period. The two-step entries are the difference between (a) the fitted value for the female wage regression (with the inverse Mills ratio) for the relevant time period and (b) the fitted value for the male wage regression for the relevant time period. The standard errors for the table entries are calculated according to the formula for the standard error of the difference between the two relevant fitted values, assuming zero correlation between them.

University of Chicago and NBER
Brown University, CEPR, and IZA

References

Aguiar, Mark, and Erik Hurst, "Measuring Trends in Leisure: The Allocation of Time over Five Decades," *Quarterly Journal of Economics*, 122 (2007), 969–1006.

Ai, Chunrong, and Edward C. Norton, "Interaction Terms in Logit and Probit Models," *Economics Letters*, 80 (2003), 123–129.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney, "Trends in U.S. Wage Inequality: Re-assessing the Revisionists," NBER Working Paper No. 11627, 2005.

Becker, Gary S., "How the Market Acted Affirmatively for Women," *Business Week*, May 13, 1985a.

——, "Human Capital, Effort, and the Sexual Division of Labor," *Journal of Labor Economics*, 3, Part 2 (1985b), S33–S58.

Bils, Mark, "Real Wages over the Business Cycle: Evidence from Panel Data," *Journal of Political Economy*, 93 (1985), 666–689.

Blau, Francine D., and Lawrence M. Kahn, "Swimming Upstream: Trends in the Gender Wage Differential in the 1980s," *Journal of Labor Economics*, 15 (1997), 1–42.

——, "Gender Differences in Pay," *Journal of Economic Perspectives*, 14 (2000), 75–99.
——, "The US Gender Pay Gap in the 1990s: Slowing Convergence," *Industrial & Labor Relations Review*, 60 (2006), 45–66.
Blundell, Richard, Amanda Gosling, Hidehiko Ichimura, and Costas Meghir, "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica*, 75 (2007), 323–363.
Borghans, Lex, Bas ter Weel, and Bruce A. Weinberg, "People People: Social Capital and the Labor Market Outcomes of Underrepresented Groups," NBER Working Paper No. 11985, 2006.
Borjas, George J., "The Economics of Immigration," *Journal of Economic Literature*, 32 (1994), 1667–1717.
Bureau of Labor Statistics, U.S. Department of Labor, *NLS of Young Women User's Guide* (Columbus, OH: Ohio State University Center for Human Resource Research, 2001).
Card, David, and John E. DiNardo, "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles," *Journal of Labor Economics*, 20 (2002), 733–783.
Chamberlain, Gary, "Asymptotic Efficiency in Semiparametric Models with Censoring," *Journal of Econometrics*, 32 (1986), 189–218.
Fortin, Nicole M., and Thomas Lemieux, "Are Women's Wage Gains Men's Losses? A Distributional Test," *American Economic Review Papers and Proceedings*, 90 (2000), 456–460.
Fuchs, Victor R., "Differences in Hourly Earnings between Men and Women," *Monthly Labor Review*, 94 (1971), 9–15.
Galor, Oded, and David N. Weil, "The Gender Gap, Fertility, and Growth," *American Economic Review*, 86 (1996), 374–387.
Goldin, Claudia, and Lawrence F. Katz, "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions," *Journal of Political Economy*, 110 (2002), 730–770.
Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko, "The Homecoming of American College Women: The Reversal of the College Gender Gap," *Journal of Economic Perspectives*, 20 (2006), 133–156.
Gottschalk, Peter, and Robert Moffitt, "The Growth of Earnings Instability in the U.S. Labor Market," *Brookings Papers on Economic Activity*, 2 (1994), 217–254.
Greene, William H., *Econometric Analysis*, 4th ed. (Upper Saddle River, NJ: Prentice-Hall, 2000).
Greenwood, Jeremy, Ananth Seshadri, and Mehmet Yorukoglu, "Engines of Liberation," *Review of Economic Studies*, 72 (2005), 109–133.
Grogger, Jeff, and Eric Eide, "Changes in College Skills and the Rise in the College Wage Premium," *Journal of Human Resources*, 30 (1995), 280–310.
Gronau, Reuben, "Wage Comparisons—A Selectivity Bias," *Journal of Political Economy*, 82 (1974), 1119–1143.
Heckman, James J., "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42 (1974), 679–694.
——, "Sample Selection Bias as a Specification Error," *Econometrica*, 47 (1979), 153–161.
——, "Varieties of Selection Bias," *American Economic Review Papers and Proceedings*, 80 (1990), 313–318.
——, "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *Journal of Political Economy*, 109 (2001), 673–748.
Heckman, James J., and Guilherme Sedlacek, "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *Journal of Political Economy*, 93 (1985), 1077–1125.
Juhn, Chinhui, and Dae Il Kim, "The Effects of Rising Female Labor Supply on Male Wages," *Journal of Labor Economics*, 17 (1999), 23–48.
Juhn, Chinhui, and Kevin M. Murphy, "Wage Inequality and Family Labor Supply," *Journal of Labor Economics*, 15, Part 1 (1997), 72–97.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce, "Accounting for the Slowdown in Black-White Convergence," in *Workers and Their Wages,* Marvin Kosters, ed. (Washington, DC: AEI Press, 1991).
——, "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, 101 (1993), 410–442.
Katz, Lawrence F., and David H. Autor, "Changes in the Wage Structure and Earnings Inequality," in *Handbook of Labor Economics, Vol. 3A* (New York: Elsevier Science, 1999), 1463–1555.
Katz, Lawrence F., and Kevin M. Murphy, "Changes in Relative Wages, 1963–1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 107 (1992), 35–78.
Keane, Michael, Robert Moffit, and David Runkle, "Real Wages over the Business Cycle: Estimating the Impact of Heterogeneity with Micro Data," *Journal of Political Economy*, 96 (1988), 1232–1266.
Levy, Frank, and Richard J. Murnane, "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," *Journal of Economic Literature*, 30 (1992), 1333–1381.
Mincer, Jacob, and Solomon Polachek, "Family Investments in Human Capital: Earnings of Women," *Journal of Political Economy*, 82, Part 2 (1974), S76–S108.
Moffitt, Robert A., "New Developments in Econometric Methods for Labor Market Analysis," in *Handbook of Labor Economics, Vol. 3A*, Orley Ashenfelter and David Card, eds. (Amsterdam: Elsevier Science/North-Holland, 1999), 1367–1397.
Mulligan, Casey B., and Yona Rubinstein, "The Closing Gender Gap as a Roy Model Illusion," NBER Working Paper No. 10892, 2004.
——, "Selection, Investment, and Women's Relative Wages since 1975," NBER Working Paper No. 11159, 2005.
Neal, Derek, "The Measured Black-White Wage Gap among Women Is Too Small," *Journal of Political Economy*, 112, Part 2 (2004), S1–S28.
Oaxaca, Ronald, "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14 (1973), 693–709.
Olivetti, Claudia, and Barbara Petrongolo, Unequal Pay or Unequal Employment? A Cross-country Analysis of Gender Gaps, manuscript, Boston University, April 2007.
O'Neill, June, "The Trend in the Male-Female Wage Gap in the United States," *Journal of Labor Economics*, 3, Part 2 (1985), S91–S116.
O'Neill, June, and Solomon Polachek, "Why the Gender Gap in Wages Narrowed in the 1980s," *Journal of Labor Economics*, 11, Part 1 (1993), 205–228.
Roy, A. D., "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3 (1951), 135–146.
Schafgans, Marcia M. A., "Ethnic Wage Differences in Malaysia: Parametric and Semiparametric Estimation of the Chinese-Malay Wage Gap," *Journal of Applied Econometrics*, 13 (1998), 481–504.
Schafgans, Marcia M. A., and Donald W. K. Andrews, "Semiparametric Estimation of the Intercept of a Sample Selection Model," *Review of Economic Studies*, 65 (1998), 497–517.
Schafgans, Marcia M. A., and Victoria Zinde-Walsh, "On Intercept Estimation in the Sample Selection Model," *Econometric Theory*, 18 (2002), 40–50.
Smith, James P., and Michael P. Ward, "Women in the Labor Market and in the Family," *Journal of Economic Perspectives*, 3 (1989), 9–23.
Solon, Gary, Robert Barsky, and Jonathan A. Parker, "Measuring the Cyclicality of Real Wages: How Important Is Composition Bias?" *Quarterly Journal of Economics*, 109 (1994), 1–25.
Topel, Robert H., "Regional Labor Markets and the Determinants of Wage Inequality," *American Economic Review Papers and Proceedings*, 84 (1994), 17–22.
Weinberg, Bruce A., "Computer Use and the Demand for Women Workers," *Industrial and Labor Relations Review*, 53 (2000), 290–308.
Welch, Finis, "Growth in Women's Relative Wages and in Inequality among Men: One Phenomenon or Two?" *American Economic Review Papers and Proceedings*, 90 (2000), 444–449.
Wellington, Alison J. "Changes in the Male-Female Wage Gap, 1976–85," *Journal of Human Resources*, 28 (1993), 383–411.