DISCUSSION PAPER SERIES

IZA

IZA DP No. 7005

Selection, Heterogeneity and the Gender Wage Gap

Cecilia Machado

November 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Selection, Heterogeneity and the Gender Wage Gap

Cecilia Machado
Getulio Vargas Foundation,
EPGE and IZA

Discussion Paper No. 7005
November 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 7005
November 2012

ABSTRACT

Selection, Heterogeneity and the Gender Wage Gap*

Selection correction methods usually make assumptions about selection itself. In the case of gender wage gap estimation, those assumptions are specially tenuous because of high female non-participation and because selection could be different in different parts of the labor market. This paper proposes an estimator for the wage gap that allows for arbitrary heterogeneity in selection. It applies to the subpopulation of "always employed" women, which is similar to men in labor force attachment. Using CPS data from 1976 to 2005, I show that the gap has narrowed substantially from a -.521 to a -.263 log wage points differential for this population.

JEL Classification: J31, J16, J24

Keywords: selection, gender wage gap

Corresponding author:

Cecilia Machado
Getulio Vargas Foundation
Graduate School of Economics
Praia de Botafogo 190, 11th floor
Botafogo, Rio de Janeiro 22250-900
Brazil
E-mail: machadoc@gmail.com

* I specially thank Douglas Almond, Janet Currie, Lena Edlund and Edward Vytlacil for encouraging support and helpful discussions, and Joshua Angrist, Tiago Berriel, David Card, Yinghua He, James Heckman, Mariesa Herrmann, Bo Honoré, Lawrence Katz, Thomas Lemieux, Dennis Kristensen, David Lee, Bentley MacLeod, Marcelo Moreira, Casey Mulligan, Derek Neal, Cristian Pop-Eleches, Ricardo Reis, Yona Rubinstein, Johannes Schmieder, Jeffrey Smith, Till von Wachter, Reed Walker, and participants at the 2009 NBER Labor Studies Fall Meeting and at the Columbia Applied Micro Colloquium, Applied Micro Workshop and Econometrics Lunch, as well as participants in seminars and conferences, for discussions and comments. All errors are my own.

1 Introduction

The narrowing of the gender wage gap in recent decades has been one of the most striking changes in the US labor market. Incidentally, over the same period, women's labor market participation increased dramatically. While many explanations for the convergence have been proposed[1], the *measurement* of wages itself is sensitive to the characteristics of the individuals who opt into employment.

On the one hand, if increased participation draws higher ability women into the labor force, gains can mechanically reflect the better characteristics of the workforce. On the other hand, if less able women choose work, the true convergence is even higher. Interestingly, both possibilites can be rationalized by economic theory.

But whether selection in fact enhances or jeopardizes women's labor market gains is subject to much dispute. In the US, while Blau and Kahn [2006] have attributted 25% of the convergence to selection[2], Mulligan and Rubinstein [2008] have found it to account for all the convergence. On the contrary, in the UK, Blundell et al. [2007] document even higher wage gains, as their selection corrected measure was higher than the observed convergence.

In common, selection correction methods assume some knowledge about the true selection mechanism. Perhaps not surprisingly, the stance taken on selection is critical in this case because female participation, albeit increasing over time, is far from full. In the US, female full-time participation still averages 50% in recent years.

When it comes to female wages and employment decisions, however, many considerations come into play. First, while positive selection – i.e., high wage individuals being more likely to work – is the gold standard in traditional labor supply models, negative selection – i.e., low wage individuals being more likely to work – is also plausible for women. If couples match based on skills and out-of-work income rises with skills, one might conjecture that

[1] Among them are the increased investments women have made in education completion and occupational choices [Goldin, 2006], changes in demand that favor women, such as increasing returns to soft-skill and reduced discrimination [Welch, 2000, Blau and Kahn, 1997], and growing cumulative labor market experience [O'Neill and Polachek, 1993]. Bailey et al. [2012] argues that the contraceptive Pill induced some of those changes, and is causaly responsible for part of the convergence.

[2] This finding is also echoed in Olivetti and Petrongolo [2008], who found wage gaps to be marginally affected by alternative correction procedures, using the same data, but for a smaller time window.

the employment decision of high skilled women reflects high reservation wages, while for low skilled women low potential earnings may be more important [Juhn and Murphy, 1997, Blundell et al., 2007, Mulligan and Rubinstein, 2008].

Second, both positive and negative selection rules can co-exist in different parts of the labor market. For instance, differences in marriage market prospects can trigger differences in selection, as women's work decision is tightly related to family structure and household income. Thus, if two groups of equally productive women face two different marriage market, one might conjecture that women under the better market are negatively selected into the workforce (and the ones who do not work have high reservation wages), whereas the women under the worse market are positively selected into work (and the ones that do not work have low market wages). Neal [2004] has found this to be the case for white and black women in the US.

And third, aside from race, several other characteristics influence marriage matches [Chiappori et al., forthcoming], and more to the point, the decision to work. Because some of those characteristics are unobserved, for whom the different positive and negative selection rules apply to is also unknown.

This paper proposes an estimator for the gender wage gap that is robust to arbitrary selection rules. In particular, it does not assume a unique selection rule (as is the case in parametric selection corrections), knowledge of the sign of average selection (which is generally assumed to tighten bounds on the wage gap) or observed selection heterogeneity (present in wage imputation procedures). In contrast to previous approaches, this estimator allows for *unobserved selection heterogeneity*, i.e. the co-existence of different selections rules in unknown parts of the labor market.

Under no prior knowledge on selection, but relying on the same type of assumptions that identify local average treatment effects – the existence of an instrument that is excluded from the wage equation and that shifts employment in the same direction for all individuals [Imbens and Angrist, 1994], – I show that average wages for a subgroup of individuals can still be identified. In analogy to the treatment effect literature [Angrist et al., 1996], I refer to this group as the "always employed", which corresponds to the individuals that always choose to work, regardless of the value of the instrument.

The instrument chosen was borrowed from Mulligan and Rubinstein [2008], and corresponds to an indicator for the presence of a child under six in the household. Thus, gender wage gaps are estimated for women that choose employment regardless of their fertility outcomes. The advantage of this instrument is twofold. First, it allows for comparability with the existing literature. And second, I argue that this always employed group of women is a relevant comparison group to men, because of their similar labor force attachments, as children induce large changes in female labor supply. Nonetheless, I also show that the failure of the identification conditions are not necessarily catastrophic for the results.

In Current Population Survey (CPS) data, I estimate an improvement in the US gender gap from 1976 to 2005. The gender wage gap is reduced by .258 log wage points, a more than twofold improvement, from -.521 to -.263 points. In the same data, I also replicate the previous findings of the literature, showing that the methods, rather than differences in the sample or in the choice of the instrument, are the driving force for the disparate results. I provide a stylized selection model featuring unobserved selection heterogeneity to show that previous estimates could deliver biased results. As female employment decisions could be arbitrarily heterogenous, the estimator proposed in this paper constitutes an alternative parameter to be incorporated in studies of the gender wage gap.

To build the case that the "always employed" women are similar to men in labor market characteristics, I turn to the National Longitudinal Survey of Youth 1979 (NLSY79), a panel data with detailed information on labor market trajectories of men and women, as well as aptitude tests, generally unobserved in more representative datasets. I mirror the "always employed" by the individuals working continuously throughout the panel, using several employment definitions. As hypothesized, men and women exhibit substantial differences in test scores, occupation choices and cumulative labor market experience over the life-cycle, but focusing on the "always employed" bridges most of those differences and indeed provides an "apples to apples" comparison between men and women.

As the"always employed" are characterized by potential actions – they would work if a child under six is present in the household and they would also work if no child under six is present – improvements in the wage gap for the "always employed" could reflect changing characteristics of this unobserved group over time. I find this not to be the case for two reasons. First, I construct a sequence of short panels in CPS, merging data for two subsequent years and identifying the individual that work on both periods. I still find convergence of the gender wage gap in that analysis, where, by definition, composition is held fixed in each panel. And second, I return to NLSY79 data and compare the "always employed" women in that cohort to the "always employed" in the Yound Women cohort of the National Logitudinal Surveys. While the fraction of "always employed" increases for the most recent cohort, I find no evidence of changing unobserved characteristic, measured by their test scores.

The baseline results were maintained under a series of robustness results such as using alternative covariate controls, accounting for male selection and redefining the instrument by different ages of the young child present. Finally, following Angrist et al. [1996], I assess the bias of the estimator when the identifying assumptions fail, and show that relaxing some of them yield qualitatively similar results.

At last, it is important to note that while the no assumption on selection is imposed, no information on the true selection mechanisms is actually recovered. The trade-off emcompasses modeling and recovering selection. In the case of the wage gap, this estimator seems appropriate because it addresses unobserved selection heterogeneity, while bridging men and women in several dimensions. But whether this estimator provides "apples to apples" in other selection applications will depend on the context.

This paper is organized as follows. The next section outlines the selection problem with missing outcomes and summarizes the existing literature. Section 3 identifies the class of models, with arbitrary heterogeneity in selection, for which a local measure of potential wages can be recovered. Section 4 presents the baseline estimates and Section 5 contains the robustness exercises. A discussion on the validity of the assumptions necessary to identify the local gender gap is found in Section 6. Section 7 concludes.

2 Backgound

2.1 The Selection Problem

Let $\mathcal{Y}$ denote the potential wages received by individuals in the market place and Y the observed wages. The selection problem arises because $Y = \mathcal{Y}$ only for those found to be employed. Since individuals who choose employment are plausibly different from the ones who do not, the observed distribution of wages does not generalize to the entire population. Thus, a simple least squares regression of observed wages on a gender dummy will mask selection effects in both the male and female population.

How big is the selection problem? Denote by $E \in \{0,1\}$ the employment status and $G \in \{0,1\}$ the gender indicator, with $G = 1$ for women. Conditional on covariates X, the unobserved mean of potential wages can be decomposed as:

$$\mathrm{E}(\mathcal{Y}|X,G) = \mathrm{E}(Y|E=1,X,G)\mathrm{Pr}(E=1|X,G) + \underbrace{\mathrm{E}(\mathcal{Y}|E=0,X,G)}_{\text{selection}}\underbrace{\mathrm{Pr}(E=0|X,G).}_{\text{non-participation}} \quad (1)$$

In words, the parameter of interest $\mathrm{E}(\mathcal{Y}|X,G)$ is a weighted average of wages among participants and non-participants. The unobservability of $\mathrm{E}(\mathcal{Y}|E=0,X,G)$ poses the main challenge in any estimation strategy, and its importance is magnified by high non-participation rates.

Without prior information on the selection effect, $\mathrm{E}(\mathcal{Y}|E=0,X,G)$ could assume a wide range of values. The most conservative approach bounds $\mathrm{E}(\mathcal{Y}|X,G)$ by considering best and worst case scenarios [Manski, 1990]. If $\mathcal{Y}$ is bounded by $\underline{\mathcal{Y}}$ and $\overline{\mathcal{Y}}$[3]:

$$\begin{aligned}\mathrm{E}(Y|E=1,X,G)\mathrm{Pr}(E=1|X,G) + \underline{\mathcal{Y}}\mathrm{Pr}(E=0|X,G)& \\ \le \mathrm{E}(\mathcal{Y}|X,G) \le& \\ \mathrm{E}(Y|E=1,X,G)\mathrm{Pr}(E=1|X,G) + \overline{\mathcal{Y}}\mathrm{Pr}(E=0|X,G).& \quad (2)\end{aligned}$$

[3] Bounded support is not a necessary condition to derive bounds on median wages, as illustrade in Appendix A.2.3.

Applying bounds to the evolution of the gender wage gap is further challenged by differences in selection and participation across genders and across time. This can be seen by letting X refer to time T, with $T \in \{t_1, t_2\}$:

$$\begin{aligned}\Delta(t_2) - \Delta(t_1) = &\Big\{\mathrm{E}(\mathcal{Y}|T = t_2, G = 1) - \mathrm{E}(\mathcal{Y}|T = t_2, G = 0)\Big\} \\ &- \Big\{\mathrm{E}(\mathcal{Y}|T = t_1, G = 1) - \mathrm{E}(\mathcal{Y}|T = t_1, G = 0)\Big\}, \quad (3)\end{aligned}$$

where each term in the expression has bounds given by (2). Even if wide (e.g., including zero), the resulting bounds from (3) highlight the magnitude of the missing data problem. Therefore, existing attempts to sign $\Delta(t_2) - \Delta(t_1)$ have imposed some knowledge about selection.

2.2 Existing Approaches and Related Literature

There are four main approaches to selection in the literature. The first approach uses information on observed covariates and restrictions motivated by economic models to impute values for the missing data [Neal and Johnson, 1996, Johnson et al., 2000, Neal, 2004, Blau and Kahn, 2006, Olivetti and Petrongolo, 2008]. For instance, if selection into employment is purely random after we control for a very detailed set of characteristics, one can match similar individuals, and impute wages for non-participants by the mean wage of participants [Neal, 2004]. Likewise, one can use observable characteristics to make assumptions on whether individuals place below or above median wages [Neal, 2004, Blau and Kahn, 2006, Olivetti and Petrongolo, 2008]. And finally, if panel data is available, one can search backward and forward in the data and proxy missing wages by the wage in the nearest wave [Blau and Kahn, 2006, Olivetti and Petrongolo, 2008]. In imputation methods, selection on unobservables is assumed away, as all missing data is filled in based on observed covariates.

A second approach argues that selection becomes negligible when participation rate are high [Chamberlain, 1986, Heckman, 1990]. Thus, gender wage gaps can be estimated in a smaller sample (also selected according to observed characteristics) where participation rates are close to one [Mulligan and Rubinstein, 2008]. To its disadvantage, those estimates likely do not generalize, as different characteristics and employment choices suggest differences on other unobserved dimensions as well [Altonji et al., 2005].

A third approach acknowledges that selection can be based on unobservables and models the self-selection process. In general, the correction procedure amounts to including an extra term in the wage equation, the control function, which is either known, as in parametric models [Gronau, 1974, Heckman, 1974], or, when unknown, estimated by semi-parametric or non-parametric methods (cf. Vella [1998]). Identification under the control function approach requires an exclusion restriction[4], that is, an instrument Z that shifts employment but is unrelated to wages.

In the US, Mulligan and Rubinstein [2008] have relied on a parametric correction method for measuring the gender wage gap, even though an extensive literature has demonstrated the sensitivity of selection models to several of their modeling assumptions. Recently, Huber and Melly [2012] have casted serious doubts on the validity of traditional sample selection models in female wage regressions, and in particular, to the empirical findings in Mulligan and Rubinstein [2008], as they imply that all explanatory variables are restricted to have the same effect at different quantiles of the outcome distribution.[5]

One last approach uses restriction motivated by economic theory to tighten the worse case scenario bounds on the gender wage gap. For example, an instrument that shifts participation tighten bounds by reducing the weight placed on the unobserved wages of non-participants [Manski, 1990]. In contrast to a parametric selection correction, for example, the bounding approach relying on an instrument imposes less restrictions as it need not model selection. However, in practical terms, informative bounds need to make further assumptions. In Blundell et al. [2007], the availability of an instrument is combined with a positive selection

[4] The exclusion restriction is not necessary in parametric models, but, in practice, identification without an instrument is weak, as the correction term is often a linear function of the variables entering the outcome equation directly [Vella, 1998].

[5] In a companion paper, Huber and Melly [2011] show that more general nonseparable sample selection models are not point identified. The identified set collapses to a point when further assumptions are imposed, such as additivity (separability) of the error term in the outcome equation and parametric specification of the copula function of the errors in the outcome and participation equation (cf. Arellano and Bonhomme [2011])

assumption and an additivity restriction in the wage equation for an empirical assessment of the gender wage gap in the UK. But where female employment and wages are concerned, such assumption may be especially tenuous. While positive selection is the norm for male selection, negative selection is also plausible for women. In addition, Neal [2004] provides evidence of different female selection rules by race. Where both positive and negative rules co-exist, the sign of average selection is also debatable.

This paper is mostly related to the last two approaches described above, since it relies on an exclusion restriction and allows selection to depend on unobservables. But in contrast to them, no assumptions on selection will need to be made. The next section outline the setting and identifying assumptions, and shows that an interesting parameter can still be recovered in that case.

3 A Local Measure of Potential Wages

3.1 Identification

This section is built upon the potential outcome notation of causal models [Rubin, 1974, Heckman, 1990]. In fact, the selection problem with missing outcomes is a particular case of treatment effect models, with the outcome being observed only for the individuals that opt into market work.

I adhere to two conventions in the literature. First, I do not explicitly model observed covariates, and all is taken to be conditional on $X = x$. Second, I abstract from general equilibrium effects, even thought they might be a relevant concern when there are big increases in participation rates.

For a binary Z, define E_1 and E_0 as the potential participation status when Z is externally set to 1 and 0 respectively. Following Angrist et al. [1996], define $(E_1 = 1, E_0 = 1)$ as the "always employed," $(E_1 = 1, E_0 = 0)$ the "employment compliers," $(E_1 = 0, E_0 = 1)$ the "employment defiers," and $(E_1 = 0, E_0 = 0)$ the "never employed." The model reads:

(AI) Existence of an Instrument:

Independence: $Z \perp\!\!\!\perp (\mathcal{Y}, E_0, E_1)$.

Nontrivial Z: $\Psi = \Pr(E=1|Z=1) - \Pr(E=1|Z=0) \neq 0$,

$$\text{with} \begin{cases} \Pr(E=1|Z=1) > 0 & \text{when } \Psi < 0 \\ \Pr(E=1|Z=0) > 0 & \text{when } \Psi > 0. \end{cases}$$

(AII) Exclusion Restriction:

$$Y = \mathcal{Y} \text{ if } E = 1,$$

$$E = ZE_1 + (1-Z)E_0.$$

(AIII) Monotonicity: Either $E_1 \leq E_0$ or $E_1 \geq E_0$ for all individuals.

Under (AI)-(AIII), mean wages for the "always employed," $\mathrm{E}(\mathcal{Y}|E_1 = 1, E_0 = 1)$, can be identified. For $E_1 \leq E_0$:

$$\begin{aligned} \mathrm{E}(Y|E=1, Z=1) &= \mathrm{E}(\mathcal{Y}|E_1 = 1, Z = 1) \\ &= \mathrm{E}(\mathcal{Y}|E_1 = 1) \\ &= \mathrm{E}(\mathcal{Y}|E_0 = 1, E_1 = 1). \end{aligned} \tag{4}$$

For $E_1 \geq E_0$, a similar reasoning shows that $\mathrm{E}(\mathcal{Y}|E_1 = 1, E_0 = 1)$ is identified by $\mathrm{E}(Y|E = 1, Z = 0)$[6]. In contrast to the treatment effect literature, which uses the instrument to identify the treatment effect among "compliers", the instrument is used here to identify potential wages for the subsample of individuals who do not change their employment decision and remain working no matter the value of Z.

Assumption (AIII) is a monotonicity restriction that rules out the existence of either $(E_0 = 0, E_1 = 1)$ or $(E_0 = 1, E_1 = 0)$ behavior[7]. Since the estimator of $\mathrm{E}(\mathcal{Y}|E_0 = 1, E_1 = 1)$

[6]Note that $Z \perp\!\!\!\perp (\mathcal{Y}, E_0, E_1)$ implies $\mathrm{E}(\mathcal{Y}|E_1 = 1, Z = 1) = \mathrm{E}(\mathcal{Y}|E_1 = 1)$ and $\mathrm{E}(\mathcal{Y}|E_0 = 1, Z = 0) = \mathrm{E}(\mathcal{Y}|E_0 = 1)$, which are weaker identification conditions. I maintain $Z \perp\!\!\!\perp (\mathcal{Y}, E_0, E_1)$ for ease of exposition.

[7]Monotonicity in selection has also been considered in treatment effect models where the outcome of interested is missing for some individuals in both treated and non-treated groups [Angrist, 1995, Zhang et al., 2008, Lee, 2009].

is sensitive to the excluded type, the direction of monotonicity should be verified in a first step. Under (AIII), this information is recovered by inspecting whether the instrument decreases or increases the employment probability, i.e., whether $\Psi = \Pr(E = 1|Z = 1) - \Pr(E = 1|Z = 0)$ is negative or positive. A negative Ψ rules out the $(E_0 = 0, E_1 = 1)$ behavior, and monotonicity holds in the decreasing direction, with $E_1 \leq E_0$. Similarly, a positive Ψ rules out the $(E_0 = 1, E_1 = 0)$ behavior, and monotonicity holds in the increasing direction, with $E_1 \geq E_0$[8].

Conditions (AI)-(AIII) also allow identification of mean wages of the "employment defiers" when $\Psi < 0$ and mean wages of the "employment compliers" when $\Psi > 0$[9]. In the empirical application that follows, I focus on the estimation of mean wages of the "always employed" women, as female employment attachment is low relative to men. This approach should yield a comparable group of men and women, which bridges differences in attachment across gender in the measurement of the wage gap.

Taken at face value, conditions (AI)-(AIII) are a subset of the assumptions used, for

[8] This can be seen by examining the expressions for $\Pr(E_0 = 0, E_1 = 1)$ and $\Pr(E_0 = 1, E_1 = 0)$. For $\gamma < 0$:

$$\begin{aligned} \Pr(E_0 = 1, E_1 = 0) &= \Pr(\alpha \geq \epsilon_i, \alpha + \gamma < \epsilon_i) \\ &= \Pr(\alpha + \gamma < \epsilon_i \leq \alpha) \\ &= F_\epsilon(\alpha) - F_\epsilon(\alpha + \gamma) \\ &= \Pr(E = 1|Z = 0) - \Pr(E = 1|Z = 1) \\ \Pr(E_0 = 0, E_1 = 1) &= \Pr(\alpha < \epsilon_i, \alpha + \gamma \geq \epsilon_i) = 0. \end{aligned}$$

Similarly, for $\gamma > 0$:

$$\begin{aligned} \Pr(E_0 = 1, E_1 = 0) &= 0 \\ \Pr(E_0 = 0, E_1 = 1) &= \Pr(E = 1|Z = 1) - \Pr(E = 1|Z = 0). \end{aligned}$$

Thus, since $\Psi = \Pr(E = 1|Z = 1) - \Pr(E = 1|Z = 0)$ is either positive or negative, either $(E_0 = 1, E_1 = 0)$ or $(E_0 = 0, E_1 = 1)$ will be assumed away by monotonicity.

[9] The estimators are given by

$$\begin{aligned} \mathrm{E}(\mathcal{Y}|E_0 = 1, E_1 = 0) &= \frac{p(0)}{p(0) - p(1)}\mathrm{E}(Y|E = 1, Z = 0) - \frac{p(1)}{p(0) - p(1)}\mathrm{E}(Y|E = 1, Z = 1) \\ \mathrm{E}(\mathcal{Y}|E_0 = 0, E_1 = 1) &= \frac{p(1)}{p(1) - p(0)}\mathrm{E}(Y|E = 1, Z = 1) - \frac{p(0)}{p(1) - p(0)}\mathrm{E}(Y|E = 1, Z = 0) \end{aligned}$$

where $p(z) = \Pr(E = 1|Z = z)$.

instance, in a parametric selection model[10]. However, those conditions alone allow for arbitrary heterogeneity in selection. In Appendix A.1 I provide a simple example of unobserved selection heterogeneity, and show that both parametric selection correction methods and a bounding approach that imposes the sign of selection fail to recover the true average or median wage. Not surprisingly, wages for the "always employed" can be recovered, as the example was taylored to satisfy conditions (AI)-(AIII). So long as those conditions are satisfied, wages for the "always employed" can be recovered also in more general models. But the benefit of recovering wages for the "always employed" using less assumptions does not come without a cost. Since selection is not modeled, the patterns of selection are also not recovered.

3.2 Estimation

The parameter of interest is the gender wage gap between the men and women of similar characteristics. Therefore, gender gaps are computed within cells of X, the vector of covariates. Abstracting from selection effects in the male population, and using Z as an instrument for female employment, a first step inspects whether

$$\begin{aligned}\Psi_{xt(G=1)} = \Pr(E = 1|Z = 1, X = x, T = t, G = 1) \\ - \Pr(E = 1|Z = 0, X = x, T = t, G = 1)\end{aligned} \quad (5)$$

is positive or negative. Under the monotonicity assumption (AIII), a negative $\Psi_{xt(G=1)}$ implies $E_1 \leq E_0$, rendering $E(Y|E = 1, Z = 1, X = x, T = t, G = 1)$ as the local estimator of women's wages. When $\Psi_{xt(G=1)}$ is positive, the estimator is $E(Y|E = 1, Z = 0, X = x, T = t, G = 1)$.

The second step is a simple OLS regression where the instrument Z enters interacted with gender:

$$Y_i = \beta_{0xt} + \beta_{1xt}G_i + \beta_{2xt}G_i * Z_i + u_i. \quad (6)$$

[10]For the equivalence between the monotonicty condition and a latent index structure, see Vytlacil [2002].

The local measure of the gap is then given by:

$$\begin{aligned}\Delta(x,t) &= \mathrm{E}(\mathcal{Y}|E_1=1,E_0=1,X=x,T=t,G=1,)-\mathrm{E}(\mathcal{Y}|X=x,T=t,G=0)\\ &= \mathrm{E}(Y|E=1,Z=1,X=x,T=t,G=1)-\mathrm{E}(Y|E=1,X=x,T=t,G=0)\\ &= \begin{cases}\beta_{1xt} & \text{if } \Psi_{xt(G=1)}>0\\ \beta_{1xt}+\beta_{2xt} & \text{if } \Psi_{xt(G=1)}<0.\end{cases}\end{aligned}$$

If male selection effects are of concern, and Z is a valid instrument for the male population as well, a first step, as in equation (5), should also be estimated for $G=0$. The second step is a variant of equation (6), where the instrument interact with both gender indicators:

$$Y_i=\beta_{0xt}+\beta_{1xt}G_i+\beta_{2xt}G_i*Z_i+\beta_{3xt}(1-G_i)*Z_i+u_i. \tag{7}$$

The local measure of the gap between "always employed" men and women is given by:

$$\begin{aligned}\Omega(x,t) &= \mathrm{E}(\mathcal{Y}|E_1=1,E_0=1,X=x,T=t,G=1)\\ &\quad -\mathrm{E}(\mathcal{Y}|E_1=1,E_0=1,X=x,T=t,G=0)\\ &= \begin{cases}\beta_{1xt} & \text{if } \Psi_{xt(G=1)}>0 \text{ and } \Psi_{xt(G=0)}>0\\ \beta_{1xt}-\beta_{3xt} & \text{if } \Psi_{xt(G=1)}>0 \text{ and } \Psi_{xt(G=0)}<0\\ \beta_{1xt}+\beta_{2xt} & \text{if } \Psi_{xt(G=1)}<0 \text{ and } \Psi_{xt(G=0)}>0\\ \beta_{1xt}+\beta_{2xt}-\beta_{3xt} & \text{if } \Psi_{xt(G=1)}<0 \text{ and } \Psi_{xt(G=0)}<0.\end{cases}\end{aligned}$$

4 The Gender Wage Gap in the US

4.1 Data

The data used in this paper comes from the Annual Demographic File (ADF) of the CPS from 1976 to 2005 and follows the sample restriction typically employed in studies of the gender gap: I focus on white non-Hispanic adults between ages of 25 and 44. The age restriction is tighter than in previous studies[11] because the instrument employed in this paper, which is fertility related, affects women of childbearing age. I define participation by two employment variables: any work and full-time-full-year work (35+ hours per week and 50 weeks or more) during the year. The outcome variable is log hourly wages. More details on the sample is found in Appendix A.3.

The instrument Z is a binary indicator for a presence of a child less than six years-old in the family. The bulk of the variation in this variable comes before age 44, as only 2% of women between 45-54 have a child younger than six years old. Moreover, although this variable is originally multivalued in the CPS survey, roughly 90% of my sample has either no children or only one child below the age of six, motivating the classification of the binary instrument. The choice of the instrument, although questionable, follows the previous literature. In Heckman [1974], one of the seminal works on female selection, number of children is used as an explanatory variable in the shadow price function. More recently, Mulligan and Rubinstein [2008] have used number of children younger than six interacted with marital status as variables determining employment, which are excluded from the market wage equation. A discussion about this instrument, and its relation to assumptions (AI)-(AIII), is found in Section 6 of this paper.

Summary statistics for the data are displayed in Table 1. Female participation increases from 65% to 80% over the period of analysis, a trend that is followed by the full-time full-year (FTFY) rate, at lower levels. Still by 2005, FTFY wages are only observed for 50% of women in the sample. The very high degree of missing wage information in the FTFY sample justifies having any employment as an alternative participation variable, bearing in mind that hourly wages for part-time workers could be smaller on average, and that the fraction of part-timers should be higher in the female population. Relative to women, male employment rates are substantially more stable, though over this period FTFY wages are not observed for more than 20% of men.

The race, ethnicity and age restrictions on the sample makes the universe of men and

[11] In Mulligan and Rubinstein [2008], the sample encompasses ages 25-54, and in Blau and Kahn [2006] it includes ages 18-65

women very similar in observables aside from two other characteristics, which are marital status and education. Although the fraction married is similar for both male and female populations, the education distribution and its evolution between 1976 and 2005 does not display a similar pattern across gender. For instance, the fraction with a college degree or more increases 5 percentage points for men, a 16% change, whereas it increases by 16 points for women, an almost twofold change. The empirical analysis that follows takes X to be education and stratifies results by 4 groups: less than high school, high school graduates, some college and college graduate or more.

4.2 Results

Since the selection problem is more severe for women – made evident by their low employment rates when compared to men – this section abstracts from male selection and takes equations (5) and (6) as benchmarks in the estimation.

Table 3 displays the first stage results stratified by the four education groups for the first and last years of the sample. The presence of a child under six decreases female employment, both for any work or full time work, with the effect being stronger at higher levels of education (where participation levels are higher). Overall, results indicate that the presence of a child younger than six decreases participation, and the sensitivity is slightly smaller for years 2001-05 relative to 1976-80.

Since $\hat{\Psi}_{xt(G=1)}$ is negative for all education groups and periods, implying monotonicity in a decreasing direction, the local measure of the gender wage gap is recovered by $\hat{\beta}_{1xt} + \hat{\beta}_{2xt}$ estimated through equation (6). Results are displayed in Table 4. Each panel of the table, one for each education group, has four regressions, which differ according to years, 1976-1980 versus 2001-2005, and to the participation classification, any employment or FTFY.

Education-wise comparison of the gender gaps indicate that they get smaller (in absolute value) as education increases, and women in the high end of the education distribution are found less subject to a penalty. Nonetheless, the local measure of the wage gap has decreased for all education groups between 1976 and 2005. The largest improvement in the gap, a twofold reduction, has occurred for the group with a college degree or more. For them, note that the local gap has closed by .20 log points, whereas the observed (or uncorrected) gap, displayed in the last line of the panel, indicates only a .10 log point reduction.

The above results can be summarized by weighting each education gap $\Delta(x,t)$ by corresponding education proportions. Since the education distribution varies over time and by gender (see Table 1), alternative weighting schemes can be employed. I consider four types of weights and display results in Table 5. The first weight, the female variable weight, uses the female education proportions in each time period, $p_{xt} = \Pr(X = x|G = 1, T = t)$, and computes the average gap by:

$$\Delta(t) = \sum_{x=1}^{4} \Delta(x,t) * p_{xt}. \tag{8}$$

The observed evolution of the gap, without selection corrections, provides a modest proxy for the gap of "always employed women": .306 versus .238 points for the ones with any employment and .258 versus .182 for the ones in full time full year work. As changes in this average gap reflect changes in each conditional gap as well as changes in the education composition of the female population, the next two weights in Table 5 hold education fixed using either its 1976-80 or its 2001-2005 proportions. The female fixed 1976-1980 weight uses $p_{xt=\text{1976-1980}}$ and the female fixed 2001-2005 weight uses $p_{xt=\text{2001-2005}}$. These alternative weighting schemes show that although part of the improvement is due to changes in the educational composition of the female workforce, the bulk of the change is due to a uniform reduction in the gender gap for each education category. Taking the education proportions in the male population as weights, $p^M_{xt} = \Pr(X = x|G = 0, T = t)$, the gains are slightly smaller and reflect the fact that the education proportions in the female population are skewed towards the groups with the highest gains.

4.3 Comparison to Previous Gender Gap Estimates

Putting the results of the previous section into perspective, Figure 1 compares $\Delta(t)$ to the observed (or "uncorrected") evolution of the gap and the gap from a parametric selection model that uses the same instrument. Appendix A.2 outlines how those two measures were obtained in my sample.

The wage gaps portrayed in the figure has participation defined as full time work and weighs the education groups by p_{xt}. The initial and final estimates of the local gap in the figure correspond to the numbers in the first line of Table 5, panel B, columns (1)-(3): an improvement of .25 log wage points, from -0.521 to -0.263. The observed gap displays a similar trend, with the measured improvement being lower, at .18 points, from -0.447 to -0.265. In contrast, the measure from a parametric selection model shows no improvement of the pay gap, which remains around -.30 points from 1976 to 2005. Note that these numbers closely track the estimates in the literature[12].

I also estimate non-parametric bounds on the median wage gap in my sample following the procedure and assumptions in Blundell et al. [2007], who have used data from the UK. Details about the computation of the bounds are also contained in Appendix A.2. Two key features are worth noting. First, bounds pertain the median, rather than the mean, wage gap, as bounds on the mean wage gap require wages to have a bounded support, which is likely not the case. Second, for purpose of comparison to Blundell et al. [2007], I maintain the results stratified by education groups and assume that the changes in the education gap is the same for all ages. I replicate their findings for the US, and find that a positive selection assumption[13] is key to determining that the relative wages of women have increased. This result can be seen in Figure A.II, which plots bounds on the changes of the gender wage gap between 1976 and 2005. A positive number indicates that gap in 2005 is lower relative to

[12] Blau and Kahn [2006] measure the observed differential as being -.459 in 1979 and -.227 in 1998 using PSID data. Mulligan and Rubinstein [2008] measure the observed differential to be -.414 in 1975-79 and -.254 in 1995-1999 using the CPS ADF data. Their parametric selection estimator of the gap is -.337 in 1975-79 and -.339 in 1995-1999.

[13] Positive selection is imposed through a stochastic dominance assumption as defined in Blundell et al. [2007].

1976, and a negative number indicates it is higher.

What do all these estimates reflect? On the one hand, selection considerations are very important for measuring wages, as female employment rates are still relatively low. This would, in principle, make the observed evolution of the gap a poor proxy for its "true" evolution. On the other hand, any attempt to correct for selection needs to impose some structure into a selection model. As would be expected, estimates using different strategies and assumptions yield conflicting answers: Mulligan and Rubinstein [2008] find that the gender wage gap has remained stable, whereas Blundell et al. [2007] find an improvement. Moreover, I am able to replicate both these findings in a single US data set illustrating that it is the method that drives the differing results, rather than any difference in sampling or choice of instrument.

The importance of unobserved heterogeneity has long been acknowledged in treatment effects models. In his Nobel lecture, James Heckman emphasized the empirical finding that "people are diverse and that diversity and heterogeneity have important implications for how we think about economic life" [Heckman, 2000]. But somewhat surprisingly, sample selection models have not incorporated heterogeneity. For women and the labor market, that consideration is relevant because selection could be different in different parts of the labor market. If that is the case, correction methods that assume some knowledge of selection itself might fail to recover truthful parameters, as exemplified in Appendix A.1.

4.4 Characterizing the "Always Employed"

The wage gap for the "always employed" women applies to a narrow, yet well-defined, subpopulation: women who do not change participation (and remain employed) in the presence of a young child. This is plausibly the group most comparable to men, as the latter have higher attachment to the labor force and seldom leave their jobs when they have children. Female labor supply, on the contrary, is very sensitive to child birth.

But who are the "always employed" and how similar are they to working men? I investigate the characteristics of those women in a longitudinal dataset, the NLSY79, where I am able to identify the ones that remain employed throughout the life-cycle. Although not exactly the same[14], the NLSY79 "always employed" women should provide a close approximation to the "always employed" in CPS because children have been shown to induce the large changes in female labor supply. Berger and Waldfogel [2004] find that approximately 63% of mothers were not working by 13 weeks after child birth, and that fraction remains high at 45% still by week 52.

Table 6 considers five alternative characterization of the "always employed" in the NLSY79 panel, considering employment as whether working any hour, any week, more than 35 hours per week, more than 50 weeks per year, and both more that 35 hours per week and more than 50 weeks per year. Appendix A.3 describes the sample delimitation and variables generated in this analysis. In all cases, the percentage of men classified as "always employed" is higher that the same percentage for women, indicating that women are less attached to the labor market.

I examine the hypothesis that the "always employed" women and men are a comparable group of individuals by looking at some important characteristics to wage determination, such as ability, occupation choice and cumulative work experience. Panel A compares all men and women in the sample, and we see the two groups are statistically different in all of them, with women being the underpriviledged group: they rank lower in the test score distribution, are less likelly to be in managerial and professional occupations, and accumulate less work experience. At the other extreme, Panel B compares the FTFY "always employed", and, as expected, we cannot reject that the two groups are similar in most characteristics, aside from cumulative experience, measured by cumulative hours, by age 44. Thus, while the FTFY "always employed" women accumulated 3,561 less work hours by age 44, that difference is substantially smaller than the 14,132 hour difference for the entire sample. Panels C to F consider lax characterization of the "always employed", and as less restrictions are imposed,

[14] Neither the panel (NLSY79) nor the cross-section (CPS) "always employed" are nested within each other. The panel group includes women that remain employed for any reason, not only children, and thus is a subset of the cross-section group. However, the panel group might also include childless women that might have chosen not to work if they had a child, and so are not "always employed" with respect to children, as defined in the cross-section.

the more different become the groups of men and women.

5 Robustness

This section examines the sensitivity of the estimates of the local measure of the gap to: covariates other than education; male selection; and alternative instruments related to the age of the young child present. For the purpose of comparison to the estimates in Figure 1, I consider participation to be full time employment, and use the female variable weights p_{xt} when averaging is necessary.

5.1 Controlling for Other Covariates

The analysis in previous sections has used education as the single covariate in X. However, selection effects may vary along other characteristics, such as age and marital status. In this section, I follow Card [1996] and Lee [2009] and incorporate all available covariates in a "skill" index. The index is used to sort workers into groups of similar characteristics and the local measure of the gender wage gap is computed within the groups.

The procedure is as following. For each period and gender, I estimate a wage equation[15] and use the model to predict wages for the entire sample, whether working or not. I then compute the four quartiles of the predicted wage distribution and sort observations into each quartile. Finally, for each period, I compute the gender gap between men and women that have the same rank on its own predicted wage distribution. I estimate equations (5) and (6) taking X to be the gender specific predicted wage quartiles. Although the predicted wage distribution varies by gender, this exercise aims to recover gaps under the assumption that the skills are approximately the same, but are rewarded differently in the market place.

Table 7 presents the summary statistics of the the sample by the four quartiles of predicted wages, stratified by period and gender. It confirms that education is an important

[15]The explanatory variables are: five education dummies (some high school, high school graduates, some college, college graduates and more than college, relative to less than high school education), three age dummies (ages 30 to 34, 35 to 39 and 40 to 44, relative to ages 25 to 29) and four marital status dummies (widowed, divorced, separated and never married, relative to married).

variable in the classification of skill types. For example, in the 1976-1980 period, the female composition in the lower quartile has education attainment below high school, whereas the upper quartile has women with at least some college education. Nonetheless, it also shows that the other covariates also have explanatory power in the skill classification, and that education is not its unique determinant. A child younger than six years old decreases female full time employment for all skill groups and time periods, as can be seen in the last two lines of panel A.

Estimates of the local gap are displayed in Figure 2. The local gender wage gap is lower, in level, for the lowest skill group, and possibly reflects minimum wage policy limiting disparities in the low end of the wage distribution. The trend towards wage equality between 1976 and 2005 is verified for all skill groups. But, most strikingly, the figure also shows continuous improvement of the gender wage gap throughout the 1990s for the two upper quartiles of the skill distribution. For purpose of comparison, Figure 3 displays the uncorrected gap by the skill types, and confirms the finding of slowing convergence of the gender gap after 1990, as documented by Blau and Kahn [2006]. Taken together, the two figures show that selection effects have masked substantial improvements in the gap in 1990s.

5.2 Male Selection

The approach taken so far has assumed that the observed wage distribution for men proxies the distribution of potential wages. In the US, however, male selection into work may challenge this assumption as one quarter of wage information is missing for the full-time employed sample, as seen in Table 1. If average selection is positive for men, and the ones participating have the highest wages, the results in previous sections constitute an exaggerated estimate of the wage gap, as the average potential wages of men should be lower.

In principle, the presence of a young child can also be used as an instrument for male participation. In fact, the summary statistics in Table 1 show that the full time employment of men increases by 9% when a young child is present. Relative to women, the sensitivity of male participation to the presence of a young child is smaller and goes in the opposite direction.

Since $\Psi_{xt(G=1)} < 0$ and $\Psi_{xt(G=0)} > 0$, the gender wage gap between "always employed" men and women is given by $\hat{\beta}_{1xt} + \hat{\beta}_{2xt}$ estimated through equation (7). Results are summarized in Figure 4. Relative to the measure of $\Delta(t)$, which only accounts for female selection, $\Omega(t)$ pictures a reduced wage gap. Male selection considerations becomes more important towards the end of the sample period, as the wedge between $\Delta(t)$ and $\Omega(t)$ gets wider. For the 2001-2005 period, the point estimates of $\Delta(t)$ and $\Omega(t)$ are -0.263 and -0.229, and depart by .034 log wage points (statistically significant at the 10% level).

5.3 Child Age and Female Participation

The presence of a child younger than six has a substantial impact on women's participation decisions. As seen in Table 1, full time employment of women with a young child is lower by .25 percentage points, almost half of the participation of women with no young child present. Since younger children require more maternal input, participation effects could differ by the age of the child. Most importantly, because the local gap recovers the pay penalty among "always employed" women, the subpopulation that does not change participation when a child younger than, say, one is present should be even more like men in terms of unobservables, such as job commitment. As the age of the youngest child decreases, and having Z as an indicator for his/her presence, the local measure of the gap should be smaller.

I investigate this hypothesis by utilizing an alternative dataset, the June CPS survey. The June CPS has a fertility supplement, available every other year, with information on the birth month and birth year of the last child. I make restrictions similar to the ones in the ADF sample, which are detailed in Appendix I. The new instruments considered, Z_j, are binary indicators of the presence of a child less than j, with $j \in \{1, ..., 6\}$, and are considered one at a time.

One shortcome of this alternative dataset is its small sample size, as wage information in the June sample is only recorded for individuals in the outgoing rotation groups (ORG).

Therefore, in this section, I will aggregate the June data into more sparse time groups, allocating approximately one quarter of total observation into four periods: 1979 to 1982, 1983 to 1987, 1988 to 1992, and 1994 to 2002. Summary statistics are presented in Table 2. The age of the youngest child alters participation in a similar manner, and the difference in participation between women with and without a child present is relatively constant (around .25 percentage points) for all values of j. In fact, full time employment can remain insensitive to the age of the child if women with young children return to part-time jobs.

Figure 5 summarizes the local gender wage gap using the alternative Z_j's as instruments. The figure suggests that as the age of the young child decreases, the local gender gender wage gap gets smaller. This result is in line with the conjecture that "always employed" women with a very young child share similar unobserved job characteristics with men, and the wage difference among them is tighter. The standard errors of the estimates, however, do not allow the inference that the local gap using Z_1 is statistically different from the one using Z_6 for the later periods in the sample.

6 Discussion

6.1 Changes in the Composition of the "Always Employed" over Time

The instrumental variable approach proposed in this paper recovers a local measure of the gender wage gap, the gap in pay between women who would choose to participate whether or not a young child is present and similar men. One criticism of using instrumental variables to recover a local parameter regards the particular and unobserved group of individuals to which the estimator refer to. In the case of the gender gap, however, that subpopulation should be the most relevant comparison to men, as, in general, women's attachment to work is lower than men's. The analysis in section 4.4 supports this view, but is restricted to only one cohort. Thus, it is possible that the "always employed" women have changed over time.

A useful exercise backs out the proportion of "employment defiers," "always employed" and "never employed" in each year. At a minimum, changing proportions of types should indicate composition effects. Noting that:

$$\Pr(E = 1|Z = 1) = \Pr(E_1 = 1) = \Pr(E_1 = 1, E_0 = 1) + \Pr(E_1 = 1, E_0 = 0)$$
$$\Pr(E = 1|Z = 0) = \Pr(E_0 = 1) = \Pr(E_1 = 1, E_0 = 1) + \Pr(E_1 = 0, E_0 = 1),$$

and using the monotonicity condition $E_1 \leq E_0$, the proportion in each group is given by:

$$\Pr(E_1 = 0, E_0 = 1) = \Pr(E = 1|Z = 0) - \Pr(E = 1|Z = 1)$$
$$\Pr(E_1 = 1, E_0 = 1) = \Pr(E = 1|Z = 1)$$
$$\Pr(E_1 = 0, E_0 = 0) = 1 - \Pr(E = 1|Z = 0).$$

From Table 1 the proportion of "always employed" averages 27.1% for the entire sample, and becomes quite sizable by 2001-05. For that last period, it reaches 34.7% of the sample. Thus, it is possible that improvements in the wage gap reflect better characteristics of the marginal women who have become "always employed." I investigate this hypothesis pursuing two strategies.

First, I inspect the characteristics of the "always employed" in two cohorts using data from NLSY79, which is representative of women ages 14-22 in 1979, and Young Women, which is representative of women ages 14-24 in 1968. In contrast to section 4.4, I focus on women only (and not on their comparison to men) because the Young Men data, the best comparison to Young Women data, contains a different questionaire on employment, conducts sparce follow-ups and interviews only until young ages. Again, the "always employed" are mirrored by the individuals working continuously throughout the life cycle, under five alternative classifications of employment. Details of sample delimitation and variables are found in Appendix A.3.

As in CPS data, the NLSY79 cohort exhibits a higher proportion of "always employed" women in all but one employment classification. For example, if we consider as employed the individuals working positive hours, the fraction women working continuously in NLSY79 is 44.2%, but correspond to only 8.0% of women in Young Women data. However, when inspecting the average test score rank of those women, table 8 finds no evidence that the higher proportion of always employed in NLSY79 are positive selected relative to Young Women. Note that some differences in occupation choices were found across the two datasets. For instance, while NLSY79 data exhibits a higher proportion of professional occupations among the "always employed" in Panel C, it exhibits a lower proportion in managerial occupations in Panel E. However, since the differences alternate in sign, this provides limited support to the hypothesis that wage gain reflect better characteristics of the "always employed" in the youngest cohort.

The second strategy identifies the "always employed" in a short CPS panel, and investigates changes in the gender wage gap for individuals employed in the two subsequent years of the data. This is possible because half of the respondents in year t are again surveyed in t+1[16]. Thus, in this panel, the composition effects are shut down by following the same individuals in the panel.

Using the short panel nature of the data (2 periods), I estimate yearly changes in the gender wage gap by:

$$Y_i = \delta_{0t} + \delta_{1t} G_i + \delta_{2t} \mathbf{1}[t+1] + \delta_{3t} G_i * \mathbf{1}[t+1] + \delta_{4t} X_i + u_i, \tag{9}$$

where t indexes the matched sample between t and t+1, and X is a vector with three education dummies (high school graduate, some college, college graduate or more, relative to less than high school). The coefficient δ_{1t} measures the gender wage gap in t, whereas δ_{3t} measures the change in the gender wage gap between t and t+1 holding fixed the population of men and women who are employed in both periods of time.

Results for δ_{3t} are displayed in Table 9, and show that improvements in the gender wage

[16]Each individual in the CPS monthly sample is eligible for eight interviews: four consecutive interviews followed by other four eight months after the fourth. Thus, individuals in their first to fourth interview in March of year t are again interviewed in March of t+1. While the match of surveys prior to 1979 is possible, the omission of identifiers makes this process tenuous, and I abstract from those years in my analysis. Neither March 1984 and March 1985, nor March 1994 and March 1995, can be merged because of revisions in household identifiers [Madrian and Lefgren, 1999]. The match rate for two subsequent periods in my sample ranges from 30% to 57%.

gap cannot be detected within each one year panel, plausibly because the interval is too short to measure gains of continuous employment. However, the compounded sum of δ_{3t} bears resemblance to changes in the gender gap for FTFY workers. Between 1979 to 2004, the change in the gender wage gap in the panel sample ($\delta_{1t=2004}$-$\delta_{1t=1979}$) was 0.206 log wage differentials, being close to the cross section measured gains for the "always employed" of 0.258. The compounded sum of δ_{3t}, while around the same order of magnitude, is not significant at conventional levels mainly due to the negative estimates of δ_{3t} in the last years of the panel. Considering the 1979-2002 time frame, the gap changes by 0.197 in the panel sample, and the compounded sum of ranges from 0.174 to 0.223, being statistically significant at the 10% level.

Convergence in the gender wage gap is verified in the panel data approach for FTFY workers, but not for part-timers as defined by any choice of hours. Thus, for the later, improvements in the gender wage gap could be reflecting changes in the pool of part-timers along the years. This exercise points to compositional changes in the extensive margin of women choosing employment, but not in the intensive margin of women choosing FTFY work. Holding fixed the composition of FTFY workers, improvements in the gender wage gap are indeed found.

6.2 The Identifying Assumptions

As in treatment effect models, solving the problem of selection with missing outcomes generally requires an instrument that shifts participation and is excluded from the outcome equation. The literature, however, has treated this assumption "quite casually," viewing identification as an intrinsic econometric problem, with little emphasis on its economic content [Moffit, 1999].

The discussion is often neglected because it is very hard to defend all the necessary assumptions for a valid instrument, particularly in a non-experimental setting. The use of the presence of a young child as an instrument in the female selection problem is no different. Although very little can be done when the instrument fails the independence condition (AI), which is the case if fertility decisions are not exogenous to employment and wages [Angrist and Evans, 1998, Browning, 1992, Willis, 1987], section 6.1 has provided additional evidence of a converging gender wage gap in the CPS panel, whose identification did not rely on an instrument.

Moreover, the failure of assumptions (AII) and (AIII) are not necessarily catastrophic. In this section, I follow Angrist et al. [1996] and discuss the sensitivity of the local measure of wages to the exclusion and monotonicity restrictions.

Monotonicity

The monotonicity condition rules out the existence of type ($E_1 = 1, E_0 = 0$) women, and would be violated if there are women who do not work, but start working once a child is present, to complement household income, for instance. By manipulation of the terms in equation (4), the bias of the local wage estimator is given by:

$$
\begin{aligned}
&\mathrm{E}(Y|E = 1, Z = 1) - \mathrm{E}(\mathcal{Y}|E_1 = 1, E_0 = 1) = \\
&\qquad\qquad \Big\{\mathrm{E}(\mathcal{Y}|E_1 = 1, E_0 = 1) - \mathrm{E}(\mathcal{Y}|E_1 = 1, E_0 = 0)\Big\}\Pr(E_0 = 0|E_1 = 1).
\end{aligned}
$$

The sign of the bias depends on whether $\mathrm{E}(\mathcal{Y}|E_1 = 1, E_0 = 1)$ is smaller or greater than $\mathrm{E}(\mathcal{Y}|E_1 = 1, E_0 = 0)$, and in principle could go either way, but should be a decreasing function of the proportion of “employment compliers” among “employment compliers” and “always employed”, $\Pr(E_0 = 0|E_1 = 1)$.

Exclusion Restriction

The exclusion restriction would fail if Z, aside from shifting employment deteriorates women’s earnings prospect. The literature on the family gap [Waldfogel, 1998] has found that women with children earn lower wages, which could reflect compositional changes among working moms combined with changes in earnings.

In this case, the model in Section 3 could be modified to:

$$Y = Z\mathcal{Y}_1 + (1 - Z)\mathcal{Y}_0 \quad \text{if} \quad E = 1$$
$$E = ZE_1 + (1 - Z)E_0,$$

where $\mathcal{Y}_1$ and $\mathcal{Y}_0$ are now the potential wages when Z is externally set to 1 and 0 respectively. Maintaining monotonicity and modifying (AIII) to $Z \perp\!\!\!\perp (\mathcal{Y}_0, \mathcal{Y}_1, E_0, E_1)$, it can be shown that $\mathrm{E}(Y|E = 1, Z = 1) = \mathrm{E}(\mathcal{Y}_1|E_1 = 1, E_0 = 1)$. That is, the instrument still recovers a local measure of wages for "always employed" when Z is set to 1. If earnings decrease in the presence of a child, $\mathrm{E}(Y|E = 1, Z = 1)$ provides a lower bound for $\mathcal{Y}$ among all "always employed."

7 Conclusion

In the early 1970s women started to enter the labor market in terms similar to men: they migrated to "high-powered" professions [Goldin and Katz, 2002] and increased their employment attachment [Goldin, 2006]. Acceptance towards the new role of women in society gained force with female rights movements, and has been corroborated by anti-discrimination laws for equal pay. Taken altogether, women have gradually converged to be as men in the workplace, looking for "careers" rather than "jobs," this being part of the larger phenomenon which Claudia Goldin termed as the "quiet revolution" that transformed women's lives.

While improvements in female workforce characteristics have been partially responsible for the closing gender wage gap, selection effects, made evident by the increase in female employment rates, could cast doubts in the relative improvement of women's wages. This paper explores the idea that likely both positive *and* negative selection rules have guided women's employment decisions. Under unobserved heterogeneity in the selection rule, I show that the usual selection correction procedures might fail to recover the parameters they target, but a local measure of the gender wage gap can still be recovered. The local gap pertains the subpopulation of "always employed" women, and is a relevant comparison group to men, as they exhibit high labor force attachment.

I conclude that the gender wage gap has indeed decreased in the US for this comparable group of men and women, and the reason for the closing gap is not due to composition changes in the "always employed" group. Empirically, I find that the observed (uncorrected) gap provides a good proxy for the local gender wage gap. I do not conclude, however, that selection effects and corrections are unimportant. On the contrary, they are so extreme in the female employment case, which is made evident by the different gap measures estimated in the literature, that a more conservative approach, as proposed in this paper, is called for. In the presence of heterogeneity in selection, focusing on the local gender wage gap is altogether less distorting than usual selection corrections.

References

Joseph G. Altonji, Todd E. Elder, and Christopher R. Taber. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1):151–184, February 2005.

Joshua D. Angrist. Conditioning on the probability of selection to control selection bias. Technical Working Paper 0181, National Bureau of Economic Research, June 1995. URL `http://www.nber.org/papers/t0181`.

Joshua D. Angrist and William Evans. Children and their parents labor supply: Evidence from exogenous variation in family size. *American Economic Review*, 88(3):450–477, June 1998.

Joshua D. Angrist, Guido D. Imbens, and Donald B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434): 444–454, June 1996.

Manuel Arellano and Stephane Bonhomme. Quantile selection models. Presentation prepared for the banco de portugal labour market conference, April 2011.

Martha Bailey, Brad Hershbein, and Amalia Miller. The opt-in revolution? contraception and the gender gap in wages. *The American economic Journal - Applied Economics*, 2012.

Lawrence Berger and Jane Waldfogel. Maternity leave and the employment of new mothers. *Journal of Population Economics*, 17:331–349, 2004.

Francine Blau and Lawrence Kahn. Swimming upstream: Trends in the gender wage differential in the 1980s. *Journal of Labor Economics*, 15(1):1–42, January 1997.

Francine Blau and Lawrence Kahn. The us gender pay gap in the 1990s: Slowing convergence. *Industrial and Labor Relations Review*, 60(1):45–66, 2006.

Richard Blundell, Amanda Gosling, Hidehiko Ichimura, and Costas Meghir. Changes in the distribution of female and male wages accounting for employment composition using bounds. *Econometrica*, 75(2):323–363, March 2007.

Martin Browning. Children and household economic behavior. *Journal of Economic Literature*, 30(3):1434–1475, September 1992.

David Card. The effect of unions on the wage structure of wages: A longitudinal analysis. *Econometrica*, 64(4):957–979, July 1996.

Gary Chamberlain. Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics*, 32:189–218, 1986.

Pierre-Andre Chiappori, Sonia Oreffice, and Climent Quintana-Domeque. Fatter attraction: Anthropometric and socioeconomic matching on the marriage market. *Journal of Political Economy*, forthcoming.

Claudia Goldin. The quiet revolution that tranformed women's employment, education, and family. *AEA Papers and Proceedings, Richard T. Ely Lecture*, 92(2):1–21, 2006.

Claudia Goldin and Lawrence F. Katz. The power of the pill: Oral contraceptives and women's career and marriage decisions. *Journal of Political Economy*, 110(4):730–770, August 2002.

Reuben Gronau. Wage comparison - a selectivity bias. *Journal of Political Economy*, 82(6): 1119–1143, 1974.

James Heckman. Shadow price, market wages, and labor supply. *Econometrica*, 42(4): 679–694, July 1974.

James Heckman. Varieties of selection bias. *American Economic Review Papers and Proceedings*, 80(2):313–318, May 1990.

James J. Heckman. Microdata, heterogeneity and the evaluation of public policy. Bank of sweden nobel memorial lecture in economic sciences, December 2000. URL http://athens.src.uchicago.edu/jenni/econ311_2002/nobel_text.pdf.

Martin Huber and Blaise Melly. Nonseparable sample selection models. mimeo, Brown University, February 2011.

Martin Huber and Blaise Melly. A test of the conditional independence assumption in sample selection models. mimeo, Brown University, September 2012.

Guido D. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, March 1994.

William Johnson, Yuichi Kitamura, and Derek Neal. Evaluating a simple method for estimating black-white gaps in median wages. *American Economic Review Papers and Proceedings*, 90(2):339–343, May 2000.

Chinhui Juhn and Kevin M. Murphy. Wage inequality and family labor supply. *Journal of Labor Economics*, 15(1):72–97, January 1997.

David Lee. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3):1071–1102, June 2009.

Brigitte C. Madrian and Lars John Lefgren. A note on longitudinally matching Current Population Survey (CPS) respondents. Working Paper 247, National Bureau of Economic Research, November 1999. URL http://www.nber.org/papers/t0247.

Charles Manski. Nonparametric bounds on treatment effects. *The American Economic Review Papers and Proceedings*, 80(2):319–323, 1990.

Robert Moffit. New developments in econometric methods for labor market analysis. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, pages 1367–1397. Elsevier, 1999.

Casey Mulligan and Yona Rubinstein. Selection, investment and women's relative wage over time. *Quarterly Journal of Economics*, 123(3):1061–1110, August 2008.

Derek Neal. The measured black-white wage gap among women is too small. *Journal of Political Economy*, 112(1):S1–S28, 2004.

Derek A. Neal and William R. Johnson. The role of pre-market factors in black-white wage differences. *Journal of Political Economy*, 104(5):869–895, October 1996.

Claudia Olivetti and Barbara Petrongolo. Unequal pay or unequal employment? a cross-country analysis of gender gaps. *Journal of Labor Economics*, 26(4):621–654, October 2008.

June O'Neill and Salomon Polachek. Why the gender gap in wages narrowed in the 1980s. *Journal of Labor Economics*, 1(1), 1993.

Donald Rubin. Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psycology*, 66(5):688–701, October 1974.

Francis Vella. Models with sample selection bias: A survey. *Journal of Human Resources*, 127-169(33):1, Winter 1998.

Edward Vytlacil. Independence, monotonicity and latent index models: An equivalence result. *Econometrica*, 70(1):331–341, January 2002.

Jane Waldfogel. Understanding the "family gap' in pay for women with children. *Journal of Economic Perspectives*, 12(1):137–156, Winter 1998.

Finis Welch. Growth in women's relative wages and in inequality among men: One phenomenon or two? *The American Economic Review, Papers and Proceedings*, 90(2):444–449, May 2000.

Robert Willis. What have we learned from the economics of the family? *American Economic Review*, 77(2):68–81, May 1987.

Junni L. Zhang, Donald B. Rubin, and Fabrizia Mealli. Evaluating the effects of job training programs on wages through principal stratification. volume 21 of *Advances in Econometrics*, pages 117–145. Emerald Group Publishing Limited, 2008.

Table 1: Summary Statistics March CPS (means)

	Female			Male		
	1976-2005	1976-80	2001-05	1976-2005	1976-80	2001-05
% married	0.709	0.791	0.671	0.666	0.776	0.616
% widowed	0.00905	0.0102	0.00873	0.00221	0.00171	0.00286
% divorced	0.114	0.0910	0.118	0.0868	0.0593	0.0916
% separated	0.0278	0.0260	0.0243	0.0192	0.0178	0.0171
% never married	0.140	0.0819	0.178	0.226	0.145	0.273
% less than hs	0.0901	0.162	0.0549	0.0960	0.147	0.0704
% hs grad	0.381	0.466	0.284	0.350	0.361	0.317
% some college	0.257	0.181	0.307	0.245	0.204	0.276
% college or more	0.273	0.191	0.354	0.309	0.288	0.336
number children ≤ 6	0.421	0.457	0.408	0.409	0.477	0.383
% no child ≤ 6	0.689	0.659	0.702	0.704	0.650	0.725
% 1 child ≤ 6	0.214	0.238	0.202	0.197	0.238	0.180
% 2 children ≤ 6	0.0853	0.0900	0.0836	0.0867	0.0987	0.0825
% 3+ children ≤ 6	0.0119	0.0125	0.0123	0.0124	0.0135	0.0122
% employed	0.763	0.648	0.795	0.946	0.965	0.931
ln(hourly wage)	1.861	1.726	2.022	2.206	2.245	2.278
% employed w/o child ≤ 6	0.822	0.737	0.841	0.934	0.955	0.915
% employed w/ child ≤ 6	0.632	0.477	0.686	0.975	0.984	0.972
% full-time, full-year	0.444	0.304	0.511	0.764	0.758	0.782
ln(hourly wage)	1.970	1.851	2.081	2.267	2.296	2.323
% FTFY w/o child ≤ 6	0.522	0.391	0.581	0.738	0.732	0.753
% FTFY w/ child ≤ 6	0.271	0.137	0.347	0.825	0.806	0.858
observations	469050	67813	92418	419208	59052	81639

Data: March CPS, ADF, 1976-2005. Sample restricted to white non-Hispanic individuals between ages 25 and 44. An individual is categorized as employed if it reports any wage in previous year. Full-time, full-year (FTFY) workers are the employed individuals who have worked 50 or more weeks, and 35 or more hours per week, in previous year. Wages are log horly wages deflated by the CPI-U, in the 1982-1984 base. More details on sample delimitation are contained in Appendix A.3. % indicates the mean of a binary variable. All means are computed using the March Supplement weight.

Table 2: Summary Statistics June CPS (means)

	Female			Male		
	1979-2002	1979-1982	1994-2002	1979-2002	1979-1982	1994-2002
% married	0.710	0.758	0.671	0.675	0.730	0.627
% absent or separated	0.0360	0.0345	0.0377	0.0249	0.0248	0.0287
% widowed or divorced	0.124	0.115	0.128	0.0859	0.0751	0.0942
% never married	0.130	0.0933	0.163	0.214	0.170	0.250
% less than hs	0.105	0.143	0.0763	0.108	0.140	0.0842
% hs grad	0.413	0.477	0.323	0.365	0.373	0.337
% some college	0.244	0.190	0.310	0.229	0.196	0.273
% college or more	0.238	0.190	0.291	0.298	0.290	0.306
% any child ≤ 6	0.325	0.339	0.324	.	.	.
% any child ≤ 5	0.279	0.291	0.274	.	.	.
% any child ≤ 4	0.233	0.244	0.228	.	.	.
% any child ≤ 3	0.180	0.190	0.177	.	.	.
% any child ≤ 2	0.124	0.130	0.122	.	.	.
% any child ≤ 1	0.0629	0.0660	0.0623	.	.	.
% full-time, full-year	0.442	0.374	0.474	0.796	0.799	0.780
ln(hourly wage)	1.928	1.871	1.980	2.195	2.224	2.184
% FTFY w/o child ≤ 6	0.522	0.458	0.551	.	.	.
% FTFY w/ child ≤ 6	0.275	0.208	0.313	.	.	.
% FTFY w/o child ≤ 5	0.509	0.445	0.537	.	.	.
% FTFY w/ child ≤ 5	0.268	0.199	0.307	.	.	.
% FTFY w/o child ≤ 4	0.497	0.432	0.526	.	.	.
% FTFY w/ child ≤ 4	0.259	0.191	0.297	.	.	.
% FTFY w/o child ≤ 3	0.484	0.419	0.513	.	.	.
% FTFY w/ child ≤ 3	0.250	0.179	0.291	.	.	.
% FTFY w/o child ≤ 2	0.471	0.405	0.502	.	.	.
% FTFY w/ child ≤ 2	0.234	0.161	0.272	.	.	.
% FTFY w/o child ≤ 1	0.456	0.389	0.488	.	.	.
% FTFY w/ child ≤ 1	0.226	0.159	0.263	.	.	.
observations	59270	14689	14697	50826	12238	13006

Data: June CPS, 1979-2002. Sample restricted to white non-Hispanic individuals between ages 25 and 44. An individual is categorized as employed if holding a job by the time of the survey. Full-time, full-year (FTFY) workers are the employed individuals who work more than 35 hours per week. Wages are log horly wages deflated by the CPI-U, in the 1982-1984 base. Age of youngest child is computed using the birth month and birth year of the child available in the fertility supplement. More details on sample delimitation are contained in Appendix A.3. % indicates the mean of a binary variable. All means are computed using weights from the earner study.

Table 3: Changes in Female Employment Induced by the Presence of Any Child ≤ 6

Entries in each triplet are: $(\Psi_{xt(G=1)}, N_{xt}, \%E_{xt})$	E: Any Employment		E: Full Time, Full Year	
	(1)	(2)	(3)	(4)
$x \backslash t$	1976-80	2001-05	1976-80	2001-05
less than high school	-0.197***	-0.133***	-0.137***	-0.162***
	10759	4831	10759	4831
	0.517	0.583	0.193	0.315
high school graduate	-0.258***	-0.150***	-0.236***	-0.215***
	31448	27025	31448	27025
	0.630	0.761	0.292	0.482
some college	-0.285***	-0.161***	-0.294***	-0.243***
	12551	29395	12551	29395
	0.680	0.812	0.322	0.514
college or more	-0.316***	-0.186***	-0.372***	-0.275***
	13055	31167	13055	31167
	0.774	0.841	0.411	0.563
pooled (all education groups)†	-0.273***	-0.166***	-0.258***	-0.244***
	67813	92418	67813	92418
	0.648	0.795	0.304	0.511

Data: March CPS, ADF, 1976-2005. Sample delimitation is as in notes to Table 1, with further details in Appendix A.3. Each triplet refers to an education (x) and time (t) specific sample. The entries are, respectively, the change in probability $\Psi_{xt(G=1)}$ as in defined in equation (5), the number of observations, and the average employment rate. The instrument Z is an indicator for the presence of child younger that six. † controls for education specific employment rates. *** denotes statistical significance at the 1% level.

Table 4: Wage Gap between Men and "Always Employed" Women, by Education

	E: Any Employment		*E*: Full Time, Full Year	
	(1)	(2)	(3)	(4)
	1976-80	2001-05	1976-80	2001-05
Panel A: Less than High School				
female, (a)	-0.572***	-0.326***	-0.502***	-0.306***
	(0.013)	(0.021)	(0.016)	(0.021)
female*Z, (b)	-0.048*	-0.058	-0.036	-0.043
	(0.026)	(0.039)	(0.038)	(0.048)
constant	2.015***	1.819***	2.063***	1.858***
	(0.008)	(0.012)	(0.008)	(0.012)
R^2	0.178	0.058	0.157	0.065
N	13298	7249	7231	4665
Local Gap, (a)+(b)	-0.620	-0.384	-0.538	-0.349
	(0.025)	(0.037)	(0.037)	(0.046)
Observed (Uncorrected) Gap	-0.582	-0.337	-0.507	-0.312
Panel B: High School Graduate				
female, (a)	-0.562***	-0.300***	-0.459***	-0.262***
	(0.007)	(0.008)	(0.007)	(0.008)
female*Z, (b)	-0.077***	-0.037***	-0.099***	-0.039**
	(0.011)	(0.013)	(0.014)	(0.016)
constant	2.225***	2.069***	2.250***	2.106***
	(0.004)	(0.005)	(0.004)	(0.004)
R^2	0.210	0.062	0.199	0.058
N	40811	45127	25420	33161
Local Gap, (a)+(b)	-0.639	-0.337	-0.557	-0.301
	(0.011)	(0.012)	(0.014)	(0.015)
Observed (Uncorrected) Gap	-0.581	-0.308	-0.475	-0.269

continued on next page

continued from previous page (Table 4)

	E: Any Employment		E: Full Time, Full Year	
	(1)	(2)	(3)	(4)
	1976-80	2001-05	1976-80	2001-05
Panel C: Some College				
female, (a)	-0.424***	-0.278***	-0.384***	-0.255***
	(0.010)	(0.007)	(0.010)	(0.008)
female*Z, (b)	-0.105***	-0.013	-0.144***	-0.027**
	(0.020)	(0.011)	(0.028)	(0.012)
constant	2.239***	2.224***	2.295***	2.262***
	(0.005)	(0.005)	(0.005)	(0.005)
R^2	0.129	0.053	0.132	0.058
Obs	20405	45847	13356	33278
Local Gap, (a)+(b)	-0.529	-0.290	-0.528	-0.282
	(0.019)	(0.011)	(0.027)	(0.012)
Observed (Uncorrected) Gap	-0.453	-0.281	-0.407	-0.261
Panel D: College or More				
female, (a)	-0.365***	-0.281***	-0.361***	-0.276***
	(0.008)	(0.008)	(0.009)	(0.008)
female*Z, (b)	-0.051***	0.109***	-0.049**	0.074***
	(0.016)	(0.010)	(0.021)	(0.012)
constant	2.382***	2.586***	2.437***	2.618***
	(0.005)	(0.005)	(0.005)	(0.005)
R^2	0.095	0.036	0.102	0.042
Obs	26844	52170	19287	39848
Local Gap, (a)+(b)	-0.416	-0.173	-0.409	-0.202
	(0.015)	(0.010)	(0.020)	(0.011)
Observed (Uncorrected) Gap	-0.378	-0.250	-0.367	-0.259

Data: March CPS, ADF, 1976-2005. Sample delimitation is as in notes to Table 1, with further details in Appendix A.3. Regression results from equation (6), stratified by education, time and participation (any employment and full time, full year). Standard errors in parenthesis. The measured gap is obtained by a wage regression that includes a gender dummy as explanatory variable. *** denotes statistical significance at the 1% level.

Table 5: Average Wage Gap, Different Weights

	Local Gap			Observed Gap		
	(1)	(2)	(3)	(4)	(5)	(6)
	1976-80	2001-05	Difference (2)-(1)	1976-80	2001-05	Difference (5)-(4)
Panel A: Any Employment						
Female Variable Weights	-0.573	-0.267	0.306	-0.519	-0.281	0.238
	(0.00787)	(0.00645)	(0.0102)	(0.00412)	(0.00398)	(0.00573)
Female Fixed Weights 76-80	-0.573	-0.305	0.268	-0.519	-0.297	0.223
	(0.00787)	(0.00879)	(0.0118)	(0.00412)	(0.00490)	(0.00640)
Female Fixed Weights 01-05	-0.525	-0.267	0.258	-0.470	-0.281	0.189
	(0.00865)	(0.00645)	(0.0108)	(0.00442)	(0.00398)	(0.00594)
Male Variable Weights	-0.549	-0.272	0.277	-0.497	-0.283	0.213
	(0.00793)	(0.00662)	(0.0103)	(0.00410)	(0.00401)	(0.00573)
Panel B: Full Time, Full Year						
Female Variable Weights	-0.521	-0.263	0.258	-0.447	-0.265	0.182
	(0.0108)	(0.00744)	(0.0131)	(0.00460)	(0.00413)	(0.00618)
Female Fixed Weights 76-80	-0.521	-0.286	0.234	-0.447	-0.273	0.175
	(0.0108)	(0.0108)	(0.0152)	(0.00460)	(0.00510)	(0.00687)
Female Fixed Weights 01-05	-0.495	-0.263	0.232	-0.418	-0.265	0.152
	(0.0119)	(0.00744)	(0.0140)	(0.00472)	(0.00413)	(0.00627)
Male Variable Weights	-0.506	-0.266	0.240	-0.435	-0.266	0.168
	(0.0109)	(0.00775)	(0.0134)	(0.00452)	(0.00416)	(0.00614)

Data: March CPS, ADF, 1976-2005. Sample delimitation is as in notes to Table 1, with further details in Appendix A.3. The table summarizes the measures of the gender wage gap from Table 4 using alternative education proportions as weights, as specified by equation (8). The female variables weights uses female education proportions in each time period. The female fixed weights holds fixed the female education proportion at either its 1976-80 or 2001-2005 level. The male variable weights uses male education proportions in each time period. Standard errors in parenthesis.

Table 6: Always Employed Men and Women, NLSY79

	A: All				B: FTFY AE			
	Men	Women	Diff	SE	Men	Women	Diff	SE
Test Score Rank	0.532	0.507	-0.024**	(0.011)	0.581	0.606	0.026	(0.036)
% Managers	0.199	0.170	-0.029**	(0.015)	0.327	0.404	0.077	(0.065)
% High Professional	0.129	0.086	-0.043***	(0.012)	0.100	0.049	-0.051	(0.039)
Total Weeks, Age 44	940	775	-165***	(8.544)	1040	1037	-2.561	(1.744)
Total Full-Time Weeks, Age 44	893	588	-305***	(10.479)	1034	1034	-0.029	(2.984)
Total Hours, Age 44	44080	29947	-14132***	(468.374)	49274	45713	-3561***	(936.237)
Observations	1316	1490	-	-	226	81	-	-
	C: FT AE				D: FY AE			
	Men	Women	Diff	SE	Men	Women	Diff	SE
IQ Rank	0.532	0.557	0.025	(0.021)	0.585	0.606	0.022	(0.030)
% Managers	0.241	0.284	0.044	(0.032)	0.304	0.348	0.044	(0.053)
% High Professional	0.091	0.102	0.011	(0.022)	0.109	0.050	-0.059*	(0.033)
Total Weeks, Age 44	1012	1006	-5	(3.607)	1034	1032	-1	(2.758)
Total FT Weeks, Age 44	1005	998	-6	(4.041)	1015	948	-66***	(12.117)
Total Hours, Age 44	48673	44293	-4380***	(555.147)	48891	43827	-5064***	(769.438)
Observations	689	280	-	-	284	121	-	-
	E: Any Hour AE				F: Any Week AE			
	Men	Women	Diff	SE	Men	Women	Diff	SE
IQ Rank	0.554	0.546	-0.008	(0.015)	0.548	0.549	0.002	(0.014)
% Managers	0.222	0.221	-0.000	(0.021)	0.221	0.210	-0.011	(0.020)
% High Professional	0.137	0.096	-0.042**	(0.017)	0.134	0.104	-0.030*	(0.016)
Total Weeks, Age 44	1003	980	-22***	(3.261)	995	971	-24***	(3.352)
Total FT Weeks, Age 44	964	806	-158***	(9.448)	954	791	-163***	(9.178)
Total Hours, Age 44	47182	39257	-7924***	(424.152)	46846	38888	-7957***	(405.076)
Observations	949	669	-	-	1057	772	-	-

Data: National Longitudinal Survey of Youth 1979 (NLSY79). Sample delimitation is described in Appendix A.3. Each panel displays average outcomes by gender, their differences (Diff) and standard errors (SE) according to the following restrictions. Panel A includes the entire sample. Panels B-F only includes individuals who are employed in each follow-up of the panel ("always employed", AE) , considering employment as: working any hour, any week, more than 35 hours per week (FT), more than 50 weeks per year (FY), and both more than 35 hours per week and more than 50 weeks per year (FTFY). * significant at 0.10; ** significant at 0.05; *** significant at 0.01.

Table 7: Summary Statistics March CPS (means), by Predicted Wage Quartile

	1976-1980				2001-2005			
	pct$\leq$25	pct 25-50	pct 50-75	pct$\geq$75	pct$\leq$25	pct 25-50	pct 50-75	pct$\geq$75
Panel A: Female								
age	31.63	36.04	31.89	33.11	31.90	37.07	35.12	37.74
% married	0.854	0.927	0.641	0.668	0.478	0.778	0.729	0.740
% never married	0.0294	0	0.131	0.204	0.252	0.0861	0.185	0.164
% less than hs	0.586	0	0	0	0.194	0	0	0
% hs grad	0.414	1	0.303	0	0.573	0.558	0	0
% some college	0	0	0.697	0.213	0.233	0.442	0.526	0
% college or more	0	0	0	0.787	0	0	0.474	1
% no child $\leq$ 6	0.549	0.732	0.630	0.719	0.666	0.746	0.722	0.677
% full time	0.211	0.253	0.364	0.426	0.455	0.501	0.551	0.543
ln(hourly wage)	1.644	1.752	1.853	2.036	1.788	1.922	2.153	2.443
% ft w/o child $\leq$ 6	0.289	0.297	0.491	0.527	0.520	0.558	0.629	0.618
% ft w/ child $\leq$ 6	0.116	0.133	0.149	0.169	0.325	0.332	0.349	0.385
Panel B: Male								
age	29.08	32.29	34.96	36.80	32.21	35.63	36.81	37.33
% married	0.544	0.747	0.878	0.957	0.330	0.676	0.663	0.892
% never married	0.355	0.153	0.0496	0.00753	0.512	0.203	0.231	0.0655
% less than hs	0.332	0.242	0.000948	0	0.226	0	0	0
% hs grad	0.481	0.395	0.510	0.00109	0.601	0.476	0	0
% some college	0.109	0.272	0.195	0.248	0.173	0.408	0.597	0
% college or more	0.0776	0.0915	0.294	0.751	0	0.116	0.403	1
% no child leq 6	0.668	0.636	0.685	0.602	0.803	0.747	0.773	0.556
% full time	0.617	0.729	0.815	0.887	0.665	0.795	0.832	0.883
ln(hourly wage)	2.056	2.215	2.327	2.531	1.984	2.217	2.398	2.727
% ft w/o child $\leq$ 6	0.578	0.699	0.800	0.879	0.641	0.774	0.820	0.865
% ft w/ child $\leq$ 6	0.695	0.782	0.849	0.899	0.767	0.856	0.875	0.906

Data: March CPS, ADF, 1976-2005. Sample delimitation is as in notes to Table 1, with further details in Appendix A.3. For each period and gender, I estimate a wage equation - the dependent variables are: five education dummies (some high school, high school graduates, some college, college graduates and more than college, relative to less than high school education), three age dummies (ages 30 to 34, 35 to 39 and 40 to 44, relative to ages 25 to 29) and four marital status dummies (widowed, divorced, separated and never married, relative to married) - and use the model to predict wages for the entire sample, whether working or not. I then compute the four quartiles of the predicted wage distribution and sort observations into each quartile. Summary statistics are presented within each quartile. % indicates the mean of a binary variable. All means are computed using the March Supplement weight.

Table 8: Always Employed Women, Young Women and NLSY79 Cohorts

	A: All				B: FTFY AE			
	Young Women	NLSY9	Diff	SE	Young Women	NLSY9	Diff	SE
Test Score Rank	0.511	0.507	-0.004	(0.011)	0.560	0.606	0.047	(0.061)
% Managers	0.179	0.170	-0.009	(0.014)	0.372	0.404	0.032	(0.094)
% High Professional	0.068	0.086	0.019*	(0.010)	0.000	0.049	0.049	(0.032)
Observations	1325	1490	-	-	34	81	-	-
	C: FT AE				D: FY AE			
	Young Women	NLSY9	Diff	SE	Young Women	NLSY9	Diff	SE
Test Score Rank	0.597	0.557	-0.040	(0.047)	0.588	0.606	0.019	(0.038)
% Managers	0.353	0.284	-0.069	(0.060)	0.354	0.348	-0.006	(0.062)
% High Professionals	0.023	0.102	0.079**	(0.036)	0.034	0.050	0.016	(0.026)
Observations	57	280	-	-	125	121	-	-
	E: Any Hour AE				F: Any Week AE			
	Young Women	NLSY9	Diff	SE	Young Women	NLSY9	Diff	SE
Test Score Rank	0.577	0.546	-0.031	(0.033)	0.560	0.549	-0.011	(0.016)
% Managers	0.289	0.221	-0.068*	(0.040)	0.237	0.210	-0.027	(0.023)
% High Professional	0.053	0.096	0.043	(0.027)	0.074	0.104	0.030*	(0.016)
Observations	97	669	-	-	497	772	-	-

Data: Young Women and National Longitudinal Survey of Youth 1979 (NLSY79). Sample delimitation is described in Appendix A.3. Each panel displays average outcomes by cohort (Young Women and NLSY79), their differences (Diff) and standard errors (SE) according to the following restrictions. Panel A includes the entire sample. Panels B-F only includes individuals who are employed in each follow-up of the panel ("always employed", AE) , considering employment as: working any hour, any week, more than 35 hours per week (FT), more than 50 weeks per year (FY), and both more than 35 hours per week and more than 50 weeks per year (FTFY). * significant at 0.10; ** significant at 0.05; *** significant at 0.01.

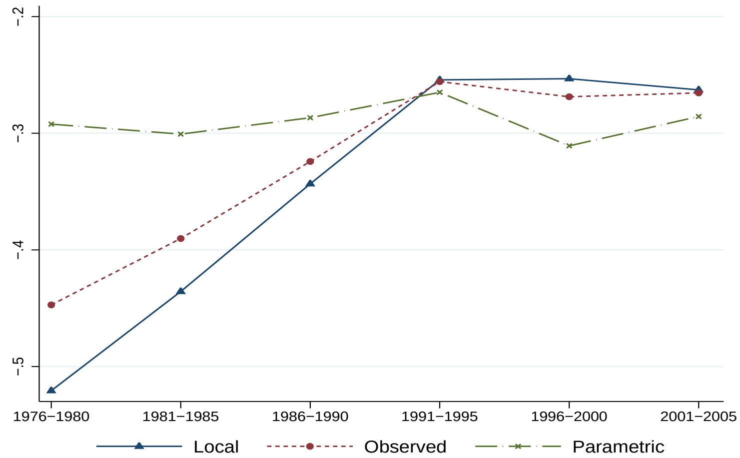
Table 9: March CPS Panel, Yearly Changes in the Gender Wage Gap

	Any Employment			Full Time		
	(1)	(2)	(3)	(4)	(5)	(6)
$t =$	N	δ_{3t}	s.e.	N	δ_{3t}	s.e.
1979	5963	0.013	(0.022)	3703	0.005	(0.023)
1980	7805	0.021	(0.020)	4821	0.027	(0.019)
1981	6872	0.022	(0.022)	4212	0.002	(0.024)
1982	7007	0.008	(0.021)	4187	0.009	(0.022)
1983	6942	0.017	(0.023)	4168	-0.000	(0.029)
1984	6915	0.012	(0.022)	4336	0.001	(0.024)
1986	7204	0.010	(0.022)	4611	0.028	(0.023)
1987	7240	0.011	(0.022)	4720	0.005	(0.023)
1988	7577	-0.009	(0.020)	5015	0.012	(0.021)
1989	9067	-0.025	(0.020)	5914	0.000	(0.020)
1990	9527	0.019	(0.018)	6189	0.024	(0.019)
1991	9376	-0.009	(0.018)	6060	0.003	(0.019)
1992	9061	-0.018	(0.019)	5891	0.006	(0.020)
1993	6751	0.008	(0.023)	4405	0.004	(0.024)
1994	5809	0.013	(0.024)	3857	0.048*	(0.028)
1996	7664	0.003	(0.021)	5166	0.003	(0.022)
1997	7231	-0.018	(0.022)	4923	-0.018	(0.023)
1998	7178	-0.006	(0.022)	4964	0.010	(0.023)
1999	7003	-0.007	(0.022)	4920	-0.035	(0.024)
2000	6559	0.014	(0.023)	4669	0.025	(0.024)
2001	6262	-0.011	(0.024)	4367	0.016	(0.027)
2002	7419	-0.036	(0.024)	5184	-0.031	(0.027)
2003	7075	-0.007	(0.025)	4850	-0.019	(0.026)
2004	6151	-0.034	(0.027)	4296	-0.036	(0.029)

	(7)	(8)	(9)	(10)
	coeff.	s.e.	coeff.	s.e.
Gap in 1979 ($\delta_{1t=1997}$)	-0.519***	(0.015)	-0.416***	(0.016)
Gap in 2002 ($\delta_{1t=2002}$)	-0.247***	(0.017)	-0.219***	(0.017)
Gap in 2004 ($\delta_{1t=2004}$)	-0.232***	(0.019)	-0.209***	(0.020)
Gap Change 2004-1979	0.287***	(0.025)	0.206***	(0.026)
Implied by $\sum_{t^a} \delta_{3t}$	0.025	(0.104)	0.125	(0.112)
Implied by $\sum_{t^b} \delta_{3t}$	0.050	(0.119)	0.173	(0.129)
Gap Change 2002-1979	0.271***	(0.022)	0.197***	(0.023)
Implied by $\sum_{t^c} \delta_{3t}$	0.067	(0.099)	0.174*	(0.105)
Implied by $\sum_{t^d} \delta_{3t}$	0.093	(0.114)	0.223*	(0.123)

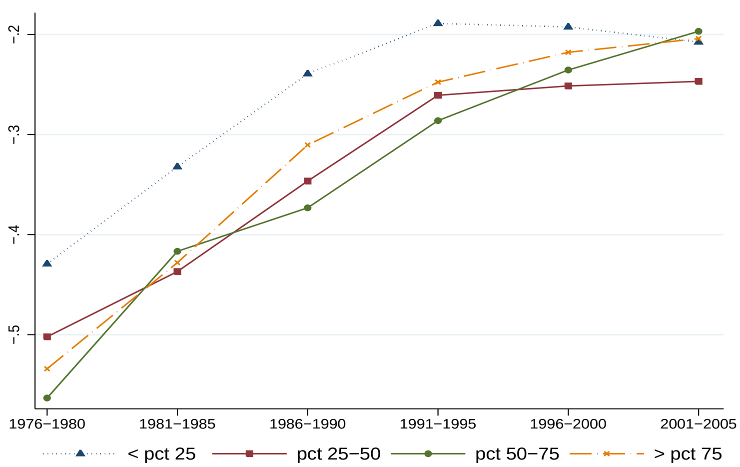
Data: March CPS, ADF, 1979-2005. Sample delimitation is as in notes to Table 1, with further details in Appendix A.3. Each line in the table refers to a two period matched sample between 1979 and 2005, where t correspond to the matched sample between t and t+1. The matching variables are the household ID, the household number, the line number and the month in sample, as described in Madrian and Lefgren [1999]. Neither March 1984 and March 1985, nor March 1994 and March 1995, can be merged because of revisions in household identifiers. The match rate ranges from 30% to 57%. For each matched sample, I estimate the change in the gender wage gap between men and women that are employed in the two subsequent periods of the data, as described by equation (9). N refers to the number of individuals in each period of the panel, δ_{3t} to the change in the gender wage gap, and s.e. is the standard error of the estimator. t^a: indexes years 1979 to 2003, except 1985 and 1995, which are the years the CPS data cannot be matched. t^b: same as t^a, but double counting years 1984 and 1994. t^c: indexes years 1979 to 2001, except 1985 and 1995. t^d: same as t^c, but double counting years 1984 and 1994. *** denotes statistical significance at the 1% level.

Figure 1: Gender Wage Gap, Local vs Observed (Uncorrected) vs Parametric



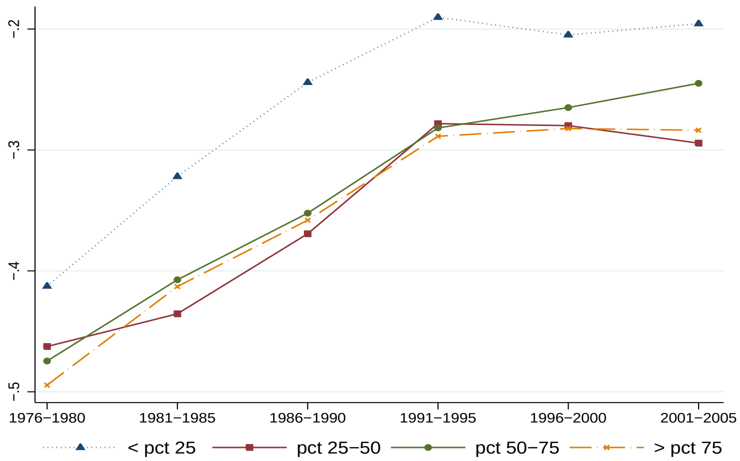
Data: March CPS, ADF, 1976-2005. Sample delimitation is as in notes to Table 1, with further details in Appendix A.3. The figure plots the gender wage gap (female minus male wages) in log wage points. The local gap averages estimates of (6), $\hat{\beta}_{1xt} + \hat{\beta}_{2xt}$, using female education proportions in each time period. The observed (uncorrected) and parametric gaps average estimates of (A.4) and (A.6), $\hat{\kappa}_{1xt}$ and $\hat{\theta}_{1xt}$, using the same weights. In all estimates, employment is defined by full time full year work. The instrument used to estimate the local and parametric gaps is an indicator for the presence of a child younger than six years old.

Figure 2: Local Gender Wage Gap, by Percentiles of Predicted Wage



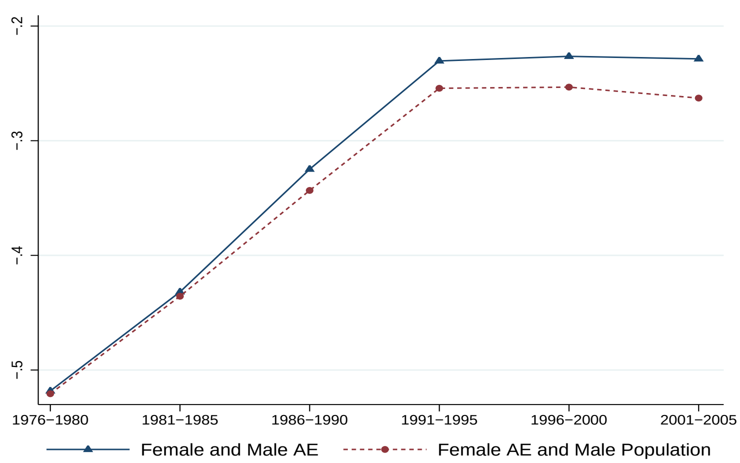
Data: March CPS, ADF, 1976-2005. Sample delimitation is as in notes to Table 1, with further details in Appendix A.3. The figure plots the local gender wage gap (female minus male wages) in log wage points, by quartiles of predicted wages. The quartiles are generated as following: for each period and gender, I estimate a wage equation - the dependent variables are: five education dummies (some high school, high school graduates, some college, college graduates and more than college, relative to less than high school education), three age dummies (ages 30 to 34, 35 to 39 and 40 to 44, relative to ages 25 to 29) and four marital status dummies (widowed, divorced, separated and never married, relative to married) - and use the model to predict wages for the entire sample, whether working or not; I then compute the four quartiles of the predicted wage distribution and sort observations into each quartile. The local gender gap is estimated between men and women that have the same rank in its own predicted wage distribution. I take X to be the four quartiles and estimate (6). In all estimates, employment is defined by full time full year work. The instrument used to estimate the local gaps is an indicator for the presence of a child younger than six years old. The lines stratify results by the four quartiles of the predicted wage distribution.

Figure 3: Observed Gender Wage Gap, by Percentiles of Predicted Wage



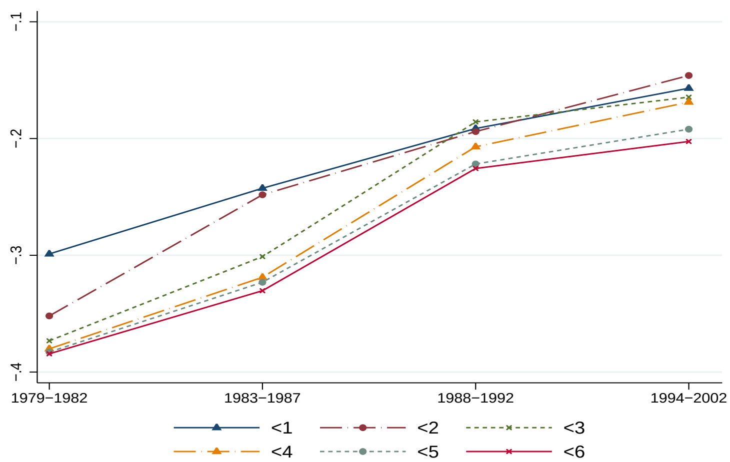
Data: March CPS, ADF, 1976-2005. Sample delimitation is as in notes to Table 1, with further details in Appendix A.3. The figure plots the observed (uncorrected) gender wage gap in log wage points, by quartiles of predicted wages. The quartiles are generated as in notes to Figure 2. The observed (uncorrected) gap is estimated by the coefficient on the female dummy in a log wage regression, as outlined in Appendix II. In all estimates, employment is defined by full time full year work. The lines stratify results by the four quartiles of the predicted wage distribution.

Figure 4: Local Gender Wage Gap, Female and Male "Always Employed"



Data: March CPS, ADF, 1976-2005. Sample delimitation is as in notes to Table 1, with further details in Appendix A.3. The figure plots the local gender wage gaps (female minus male wages) in log wage points. The "Female and Male AE" line averages estimates of (7), $\hat{\beta}_{1xt} + \hat{\beta}_{2xt}$, using female education proportions in each time period. The "Female AE and Male Population" line averages estimates of (6), $\hat{\beta}_{1xt} + \hat{\beta}_{2xt}$, using the same weights. Employment is defined by full time full year work in both lines. The instrument used to estimate the local gaps is an indicator for the presence of a child younger than six years old.

Figure 5: Local Gender Wage Gap, by Age of Young Child



Data: June CPS, 1979-2002. Sample delimitation is as in notes to Table 2, with further details in Appendix A.3. The figure plots the local gender wage gap (female minus male wages) in log wage points. The local gap averages estimates of (6), $\hat{\beta}_{1xt} + \hat{\beta}_{2xt}$, using female education proportions in each time period. In all estimates, employment is defined by full time full year work. Lines differ with respect to instrument used in estimation: child younger than j, with $j \in \{1, ..., 6\}$.

Appendix

A.1 Unobserved Selection Heterogeneity: Example

Let female employment and wages be generated by the model:

$$
\begin{aligned}
E_i &= \mathbf{1}[\alpha + \gamma Z_i \geq \epsilon_i] \\
Y_i &= \mathcal{Y}_i \ \text{ if } \ E_i = 1 \\
Z_i &\perp\!\!\!\perp (\mathcal{Y}_i, \epsilon_i),
\end{aligned}
\tag{A.1}
$$

with wages specified by the rule:

$$
\mathcal{Y}_i = \begin{cases} (\rho_{\mathtt{N}} * \sigma_{\mathtt{N}}/\sigma_\epsilon) * \epsilon_i + \zeta_i & \text{if } \epsilon_i \geq 0 \text{ (type } \mathtt{N}\text{)} \\ (\rho_{\mathtt{P}} * \sigma_{\mathtt{P}}/\sigma_\epsilon) * \epsilon_i + \zeta_i & \text{if } \epsilon_i < 0 \text{ (type } \mathtt{P}\text{)}, \end{cases}
\tag{A.2}
$$

and parametrization and data generating process (DGP) given by:

$$
\begin{aligned}
\epsilon_i &\sim N(0, \sigma_\epsilon^2), \quad \sigma_\epsilon = 4 \\
Z_i &\sim \text{Bernoulli}(0.5) \\
\alpha &= 2.5, \quad \gamma = -5 \\
\zeta_i &\sim N(0, 1) \\
(\rho_{\mathtt{N}}, \sigma_{\mathtt{N}}) &= (0.5, 4) \\
(\rho_{\mathtt{P}}, \sigma_{\mathtt{P}}) &= (-0.5, 4).
\end{aligned}
\tag{A.3}
$$

The model in (A.1) is a standard labor supply model where $\mathcal{Y}_i$ and ϵ_i are the unobservables that jointly determine employment and wages, which are generally correlated. In the terminology of those models, ϵ_i corresponds to the difference in the unobservables of the reservation and market wage equations[17]. Under a joint normality assumption, the correla-

[17] This can be seen by letting the reservation wages Y_i^R be given by $Y_i^R = -\alpha - \gamma Z_i + \xi_i$. Thus, $E_i = 1$ if $\mathcal{Y}_i \geq Y_i^R$ or, alternatively, $\alpha + \gamma Z_i \geq \xi_i - \mathcal{Y}_i = \epsilon_i$.

tion between $\mathcal{Y}_i$ and ϵ_i determines the sign of selection: if the correlation is positive, selection is said to be negative, and vice-versa[18].

Heterogeneity in the selection rule is captured by the co-existence of two rules, $\texttt{N}$ and $\texttt{P}$, corresponding to a negative and a positive selection mechanism, respectively, as seen in (A.2). This heterogeneity is unobserved because it depends on ϵ_i, being of type $\texttt{N}$ if $\epsilon_i \geq 0$ and of type $\texttt{P}$ if $\epsilon_i < 0$. Since $\epsilon_i \sim N(0, \sigma_\epsilon^2)$, the two types are equally likely. If $\rho_{\texttt{N}} = \rho_{\texttt{P}}$, we return to the conventional setup.

Table A.I displays summary statistics for 1,000 datasets, with 10,000 observations each, generated by the above model and DGP. Panel A reports some statistics on wages and employment that would be observed by researchers. Mean wages among the employed are 1.596, and are based on information about the 50% of the population that chooses to participate. The instrument Z decreases employment by 0.469 percentange points (0.734-0.265).

Panel B displays information on $\mathcal{Y}$ by the underlying selection rule, which could only be recovered under knowledge of the true model. In this example, the positive and negative rules roughly cancel each other out, and the average of potential wages is very close to the observed mean wages among the employed. Because the two selection rules are symmetric, type $\texttt{N}$ and $\texttt{P}$ women have the same distribution of wages. This can be seen in Figure A.I, which plots the distribution of wages for one of the 1,000 datasets. The two types only differ with respect to ϵ_i, and, consequently, on the likelihood of employment. Individuals with selection rule $\texttt{N}$, are less likely to be employed than type $\texttt{P}$, as higher values of ϵ_i meet the the employment threshold equation less frequently. As would be expected, type $\texttt{N}$ individuals are negatively selected into employment, with $\mathrm{E}(\mathcal{Y}|E = 0, \texttt{N}) \geq \mathrm{E}(\mathcal{Y}|E = 1, \texttt{N})$, as $\rho_{\texttt{N}} > 0$. Type $\texttt{P}$ individuals are positively selected.

This model captures the idea that the selection rule will differ according to how well a women will fare in terms of reservation wages (relative to market wages), which is what ϵ_i represents. Women with $\epsilon_i \geq 0$ have their employment decision strongly guided by high reservation wages, and among them, the ones that work have the lowest market wages

[18] When comparing $\mathcal{Y}_i$ and $-\epsilon_i$ - the difference between market and reservation wages - the sign of the correlation and selection go on the same direction.

(negative selection). The symmetric reasoning holds for women with $\epsilon_i < 0$. A microfounded model yielding these features has been proposed by Neal [2004], where selection heterogeneity arises from differences in marriage market prospects.

Panel C presents estimates on $\mathcal{Y}$ under two common approaches used to recover female wages (and the gender gap): a parametric selection model that assumes joint normality of unobservables and a unique selection rule, as in Mulligan and Rubinstein [2008], and non-parametric bounds on mean and median wages, assuming the sign of the average selection is known, as in Blundell et al. [2007].

The parametric selection model includes the Mills ratio as a control in the wage equation, and estimates the mean wages to be 0.605, which deflates its true measure of 1.595 by more than half. Based on that estimation framework, one would conclude that selection is positive, as wages of the employed, 1.596, are higher than the estimated measure of $\mathrm{E}(\mathcal{Y})$, even though positive selection is no more likely than negative selection.

Placing non-parametric bounds on mean and median wages combines the availability of the instrument Z with the assumption of either a positive or negative average selection mechanism. Information on the construction of the bounds is contained in Appendix A.2.3. Since the employment rate in the example is low, best and worse case bounds on $\mathrm{E}(\mathcal{Y})$ are very wide, and range from a low of 0.119 to a high of 3.229. Imposing average positive selection reduces the upper bound to 1.280, and the resulting bounds range from 0.119 to 1.280. Similarly, under negative selection the lower bound becomes 2.468, with the upper bound of 3.229 being maintained. Bounds under either positive or negative selection assumptions also miss $\mathrm{E}(\mathcal{Y})$, as average selection in this example is zero. With respect to median median wages, bounds are relatively tighter, but still range from 0.501 to 1.825. As in the case of mean wages, the positive selection assumption reduces upper bound to 1.129, and misses median wages, which is 1.465 in this example. The negative selection assumption is rejected in this case, as upper and lower limits cross.

This example shows that both the parametric correction and the non-parametric bounds that sign the average selection may fail to recover mean (or median) potential wages under unobserved heterogeneity in the selection rule. However, the example given by equations

(A.1) to (A.3) is one among other possible models of female employment selection. Below I examine the robustness of these results to some alternative modeling strategies.

Different Proportion of Types How do the parametric correction and non-parametric bounds fare as the degree of selection heterogeneity diminishes? In the example, type N and P are equally likely, but different proportion of types can be generated by considering different ϵ_i thresholds in equation (A.2). The parametric correction fails even for small heterogeneity (eg, 10% or 90% type P), and always conclude positive selection. As for bounds on median wages, the tolerance is slighlty higher, but for any proportion of P type individuals between 20% and 80%, negative selection would be rejected and bounds under positive selection would miss median wages.

Symmetric Assignment of Types Opposite results are obtained when modifying equation (A.2) to let $\epsilon_i \geq 0$ denote the individual under a type P rule and $\epsilon_i < 0$ the ones under a type N rule. The parametric correction concludes negative selection, whereas bounds on median wages reject the positive selection assumption and do not recover median wages under the negative selection assumption.

Equal Participation Rates Are the bias in estimates due to differences in employment participation across the types? For example, is positive selection concluded because work participation is higher among the P type? Is the negative selection assumption rejected for the same reason under the bounding approach? The answer is yes for a selection model in line with equations (A.1) to (A.3). Equal proportion of types and equal participation rates across them can be generated by letting $(\alpha_{\texttt{N}}, \gamma_{\texttt{N}}) = (4, -2)$ and $(\alpha_{\texttt{P}}, \gamma_{\texttt{P}}) = (-1, -3)$ in the employment threshold equation, for instance. The parametric correction gets close to mean wages, and neither positive nor negative selection can be rejected under the bounding approach.

Finally, panels A and D of Table A.I show that $\mathrm{E}(Y|E = 1, Z = 1)$ matches $\mathrm{E}(\mathcal{Y}|E_1 = 1, E_0 = 1)$. This is the case because assumptions (AI)-(AIII) outlined in section 3.1 are satisfied in this example, with $\Psi < 0$.

A.2 Gender Wage Gap Measures

This Appendix describes the other estimates of the gender gap pictured in Figures 1 and A.II.

A.2.1 Observed Gap

The observed gap was obtained from the following regression:

$$Y_i = \kappa_{0xt} + \kappa_{1xt} * G_i + u_i, \tag{A.4}$$

where x again subscripts four education categories: less than high school (includes some high school), high school graduate, some college, and college graduate or more. κ_{1xt} correspond to education specific gaps, and the observed gap in period t is recovered by weighting κ_{1xt} by female education proportions in each time period.

A.2.2 Parametric Correction

The gap obtained from the parametric selection model is estimated in two stages. The first stage estimates a participation probit for the female population:

$$\Pr(E = 1|T = t, X, G = 1) = \Phi(\alpha_t' X_i + \gamma_t Z_i), \tag{A.5}$$

where X_i is a vector with education dummies and Z_i indicates the presence of a child less than 6. The above results is used to construct the mills ratio, $\lambda_i = \lambda(\hat{\alpha}_t X_i + \hat{\gamma}_t Z_i)$, and the second stage regression includes $\lambda_i * G_i$ as a control for the female selection bias:

$$Y_i = \theta_{0xt} + \theta_{1xt} * G_i + \theta_{2xt}\lambda_i * G_i + u_i. \tag{A.6}$$

θ_{1xt} is weighted by female education proportions in each time period.

A.2.3 Bounds on Median Wage

Noting that the CDF of wages, $F(\mathcal{Y})$, ranges from 0 and 1, worse case bounds on $F(\mathcal{Y})$ are given by:

$$F(Y|E = 1)\Pr(E = 1) \leq F(\mathcal{Y}) \leq F(Y|E = 1)\Pr(E = 1) + \Pr(E = 0). \quad \text{(A.7)}$$

If an instrument is available, so that $F(\mathcal{Y}) = F(\mathcal{Y}|Z)$, the above expression is tightened to:

$$\max{}_Z\left\{F(Y|E = 1, Z)\Pr(E = 1|Z)\right\} \leq F(\mathcal{Y}) \leq \min{}_Z\left\{F(Y|E = 1, Z)\Pr(E = 1|Z) + \Pr(E = 0|Z)\right\}. \quad \text{(A.8)}$$

Alternatively, if positive selection is instead assumed, then wages for those observed working stochastically dominate those on non workers and $F(Y|E = 1) \leq F(\mathcal{Y}|E = 0)$. Thus, lower bound in (A.7) is tightened to $F(Y|E = 1)$. Similarly, under negative selection upper bound in (A.7) becomes $F(Y|E = 1)$.

Combining both the availability of an instrument and the sign of average selection increases lower bound in (A.8) to $\max_Z F(Y|E = 1, Z)$ in the case of positive selection and reduces upper bound in (A.8) to $\min_Z F(Y|E = 1, Z)$ in the case of negative selection.

To illustrate how the bounds on median wages are constructed, take the case given by (A.8). Let Z' (Z'') denote the argmax (argmin) of above lower (upper) bound. Then, median wages are bounded by:

$$Y^l \leq \text{Med}(\mathcal{Y}) \leq Y^u, \quad \text{(A.9)}$$

where Y^l and Y^u solve the q^l and q^u quantiles of $F(.|E = 1, Z'')$ and $F(.|E = 1, Z')$ with:

$$q^l = \frac{0.5 - \Pr(E = 0|Z'')}{\Pr(E = 1|Z'')} \quad \text{and} \quad q^u = \frac{0.5}{\Pr(E = 1|Z')} \quad . \quad \text{(A.10)}$$

Each quantity in (A.8) and (A.10) is estimated by its sample counterpart, and the observed distribution of $Y|E=1,Z$ is used when searching for Y^l and Y^u. Note that the quantiles q^l and q^u can only be computed when participation is above 50%. For example, estimation of (A.10) requires that $\Pr(E=0|Z'') \leq 0.5$ and that $\Pr(E=1|Z') \geq 0.5$.

Empirically, I follow the same approach in Blundell et al. [2007] and construct bounds on the median gender wage gap using US data, with sample delimitation for the March CPS data as described in Appendix A.3. Participation is defined by any employment[19], and conditioning covariates are time T, education X and age A. I let T denote the start and end of my sample, t_1=1976-1980 and t_2=2001-2005, X, the four education categories and A, five year age groups (25-29, 30-34, 35-39, 40-44). The parameter of interest is the improvement of the gender gap between 1976-1980 and 2001-2005, and is given by:

$$\begin{aligned}\Delta(t_2,X,A)-\Delta(t_1,X,A) = &\Big\{\mathrm{Med}(\mathcal{Y}|t_2,X,A,G=1)-\mathrm{Med}(\mathcal{Y}|t_2,X,A,G=0)\Big\}\\ &-\Big\{\mathrm{Med}(\mathcal{Y}|t_1,X,A,G=1)-\mathrm{Med}(\mathcal{Y}|t_1,X,A,G=0)\Big\}. \quad \text{(A.11)}\end{aligned}$$

Bounds on each quantity of (A.11) consider four different scenarios. The baseline result are bounds on $\Delta(t_2,X,A)-\Delta(t_1,X,A)$ that impose neither the exclusion restriction nor the sign of average selection. The second scenario imposes positive selection for both men and women, and for both time periods. The third uses the exclusion restriction, the presence of a child less than six years old, in the computation of female median wages, with no restrictions on the sign of average selection for both men and women. Finally, the fourth case combines the exclusion restriction for women with positive selection for both genders.

Results on $\Delta(t_2,X,A)-\Delta(t_1,X,A)$ are summarized by education categories using the "additivity" assumption of Blundell et al. [2007], in which changes in education differentials across time are assumed to be the same for all age groups, making it valid to search across

[19] Full time work rate is below 50% for women in start of sample period, and bounds on $\mathrm{Med}(\mathcal{Y})$ cannot be recovered.

A for the best values of $\Delta(t_2, X, A) - \Delta(t_1, X, A)$. That is:

$$\max_A LB\left\{\Delta(t_2, X, A) - \Delta(t_1, X, A)\right\} \leq \Delta(t_2, X) - \Delta(t_1, X) \leq \min_A UB\left\{\Delta(t_2, X, A) - \Delta(t_1, X, A)\right\} \quad (A.12)$$

where LB and UB stand for lower and upper bounds. Results are displayed in Figure A.II. The baseline case is uninformative about the evolution of the gap for all education categories. For those with some college, for example, the gender gap could have improved as much as .52 log points or deteriorated by 0.21 log points. The positive selection assumption is shown to be quite crucial to sign the evolution of the gap: all individuals with high school degrees or more experience a substantial improvement in gap under this assumption. The exclusion restriction alone is barely able to sign changes in gap (only for individuals with a college degree or more), even when abstracting from sample variability. Finally, the combination of the exclusion restriction, positive selection and additivity is not valid, as bounds for those with some college and more than college cross.

A.3 Data

March CPS

The Annual Demographic File (ADF) is an annual supplement conducted every March in the CPS survey. It surveys all individuals in the monthly sample, and has detailed demographic information, as well as income and labor force participation in the previous calendar year.

In this paper, I use data from 1976 to 2005. The analysis starts in 1976 because hours worked per week is a categorical variable prior to 1976. I restrict the sample to non-Hispanic whites between ages 25-44, and exclude individuals living in group quarters and the self-employed. I further eliminate observations with inconsistent reports of weeks worked per year, hours per week, and annual wages, as well as any allocation for these variables. Hourly wages are generated by dividing annual wages by annual hours (week per year times hours per week), and are deflated by the CPI-U (Consumer Price Index, All Urban Consumers) in the 1982-1984 base. The wage measure used in the paper is the log of hourly wages.

An individual is classified as employed if hourly wage information is available. The full time full year employment sample is further restricted to individuals working more that 35 hours per week and more than 50 week during the year.

Presence of a child younger than 6 is generated from "number of children in family under 6" and is available for both female and male individuals.

The other demographic information in the sample are marital status and education. I classify individuals into five marital status categories - married, widowed, divorced, separated and never married - and four education groups - less than high school (includes some high school), high school graduate, some college, and college graduate or more.

All estimates using the March CPS data employ the March Supplement weight.

June CPS

The Marriage and Fertility Supplement is conducted on the month of June of the CPS survey in selected years. The questions asked in this supplement vary by survey, but the main fertility variable used in this paper, the year and month of birth of the youngest child, is asked in every fertility supplement, and is consistently available for all females of childbearing age (18-44). I use year and month of birth of youngest child to generate indicator variables of presence of young child less than x, with $x \in \{1, ..., 6\}$.

I follow the same sample delimitation as in the March sample, keeping the non-Hispanic whites between ages 25-44 and excluding individuals living in group quarters and the self-employed. Wages and employment come from the earner study, which is conducted every month in the CPS sample for the individuals in the outgoing rotation groups (ORG) 4 and 8. Therefore, relative to the ADF file, the sample in the ORG corresponds to one fourth of monthly sample. The earner study has information on weekly wages and hours worked on the jobs individuals hold by the time of the survey. I eliminate individuals with inconsistent reports of weekly earnings and hours per week, as well as any allocation for these two variables. I generate hourly wages by the ratio of weekly earnings and hours per week, and deflate it by the CPI-U. The wage measure is log of hourly wages.

Employment is defined by individuals with hourly wage informations. The full time sample corresponds to those working more than 35 hours per week. Other demographics in the sample are marital status and education. The marital status classification is slightly different from the one used in the ADF sample, due to questionnaire differences, and has four categories: married, spouse absent or separated, widowed or divorced and never married. The education classification is the same as in the ADF sample: less than high school (includes some high school), high school graduate, some college, and college graduate or more.

Since the earner study starts in 1979, this is the first year in my June CPS sample. Survey years included in the sample follow the availability of the Marriage and Fertility Supplement and are 1979-88, 1990, 1992, 1994, 1995, 1998, 2000 and 2002.

All estimates using the June CPS data employ weights from the earner study.

National Longitudinal Surveys

The National Longitudinal Surveys (NLS) are a family of surveys dedicated to tracking labor market and other life experiences of American men and women. In this paper, I analyse two cohorts, the Young Women cohort and the National Longitudinal Survey of Youth 1979 (NLSY79) cohort.

The Young Women cohort is a representative sample of women ages 14-24 in 1968. Follow-up surveys occured in 1969, 1970, 1971, 1972, 1973, 1975, 1977, 1978, 1980, 1982, 1983, 1985, 1987, 1988, 1991, 1993, 1995, 1997, 1999, 2001 and 2003. I restrict sample to white women (hispanic information is not available at baseline) and use the panel weights provided in the data in all estimates.

In each survey year, I consider five employment definitions: whether working any hour, whether working any week, whether working more than 35 hours per week (full-time), whether working more than 50 weeks (full-year), and whether full-time and full-year. The questions on hours differ across survey years, and either refer to the hours worked per week in the current or last job or to hours usually worked per week. Weeks worked correspond to the weeks worked in the last year, and this information is not available in 1971, 1972 and 1973. The "always employed" women in this data are characterized by being employed, according to each of the five employment measures above, in every year of the follow-up surveys when between ages 25-43.

Various aptitude and intelligence scores were collected during the 1968 school survey year. I use the composite score, referred to as IQ score, designed to unify those measures across the sample and made available by the survey. For purpose of comparability with the test scores in NLSY79, I transform this IQ measure, which ranges from 40 to 160, into the rank each indivividual gets in the distribution of scores, ranging from 0 to 1. Managerial and professional occupations were classified according to the 1980 Census occupation codes, considering the most recent occupation reported in the follow-up surveys. High professionals correspond to professional occupations excluding health assessment and treating occupations, therapists, all teachers, except post-secondary.

The NLSY79 cohort is a representative sample of men and women ages 14-22 in 1979. Follow-up surveys ocurred in 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1996, 1998, 2000, 2002, 2004, 2006 and 2008. I restrict sample to white non-hispanic individuals and use the panel weights provided in the data in all estimates.

I also consider five employment definitions: whether working any hour, whether working any week, whether working more than 35 hours per week (full-time), whether working more than 50 weeks (full-year), and whether full-time and full-year. In each survey year, hours per week and weeks per year are measured in the last calendar year. The "always employed" women in this data are characterized by being employed, according to each of the five employment measures above, in every year of the follow-up surveys when between ages 25-43.

The aptitude test in NLSY79 corresponds to the respondents' AFQT scores calculated from the ASVAB tests for the vast majority of respondents who took them in 1980. The scores have been renormed twice based on updated standards. For purpose of comparability with the test scores in Young Women data, I transform this AFQT score, which ranges from 0 to 100, into the rank each indivividual gets in the distribution of scores, ranging from 0 to 1. Managerial and professional occupations were classified according to the 1980 Census occupation codes, considering the most recent occupation reported in the follow-up surveys. As in Young Women data, high professionals correspond to professional occupations excluding health assessment and treating occupations, therapists, all teachers, except post-secondary.

Table A.I: Simulated Data

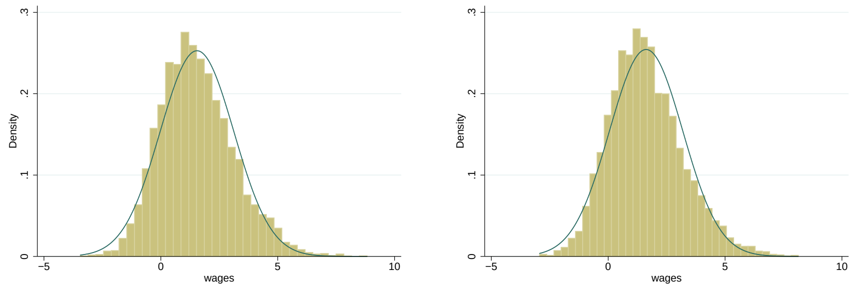
	Mean	90% CI
Panel A: Summary of Observed Data		
$\mathrm{E}(Y\|E=1)$	1.596	(1.560,1.630)
$\mathrm{E}(Y\|E=1,Z=1)$	2.468	(2.404,2.528)
$\mathrm{E}(Y\|E=1,Z=0)$	1.280	(1.241,1.320)
$\Pr(E=1)$	0.500	(0.492,0.507)
$\Pr(E=1\|Z=1)$	0.265	(0.256,0.275)
$\Pr(E=1\|Z=0)$	0.734	(0.724,0.745)
Panel B: Summary by Type		
$\mathrm{E}(\mathcal{Y})$	1.595	(1.569,1.623)
$\mathrm{Med}(\mathcal{Y})$	1.465	(1.432,1.496)
$\mathrm{E}(\mathcal{Y}\|\mathtt{N})$	1.595	(1.559,1.633)
$\mathrm{E}(\mathcal{Y}\|\mathtt{P})$	1.595	(1.560,1.632)
$\Pr(E=1\|\mathtt{N})$	0.234	(0.224,0.244)
$\Pr(E=1\|\mathtt{P})$	0.765	(0.756,0.775)
$\mathrm{E}(\mathcal{Y}\|E=1,\mathtt{N})$	0.605	(0.552,0.657)
$\mathrm{E}(\mathcal{Y}\|E=0,\mathtt{N})$	1.898	(1.855,1.940)
$\mathrm{E}(\mathcal{Y}\|E=1,\mathtt{P})$	1.899	(1.858,1.941)
$\mathrm{E}(\mathcal{Y}\|E=0,\mathtt{P})$	0.604	(0.551,0.657)
Panel C: Summary by Potential Response		
$\mathrm{E}(\mathcal{Y}\|E_1=0,E_0=0)$	2.467	(2.422,2.510)
$\mathrm{E}(\mathcal{Y}\|E_1=0,E_0=1)$	0.605	(0.579,0.629)
$\mathrm{E}(\mathcal{Y}\|E_1=1,E_0=1)$	2.467	(2.424,2.510)
$\Pr(E_1=0,E_0=0)$	0.266	(0.258,0.274)
$\Pr(E_1=1,E_0=0)$	0.468	(0.460,0.477)
$\Pr(E_1=1,E_0=1)$	0.266	(0.259,0.273)
Panel D: Estimation		
Parametric Estimator of $\mathrm{E}(\mathcal{Y})$[a]	0.605	(0.518,0.688)
LB on $\mathrm{E}(\mathcal{Y})$	0.119	(-0.060,0.261)
UB on $\mathrm{E}(\mathcal{Y})$	3.229	(2.966,3.571)
UB on $\mathrm{E}(\mathcal{Y})$, Positive Selection[b]	1.280	(1.241,1.320)
LB on $\mathrm{E}(\mathcal{Y})$, Negative Selection[c]	2.468	(2.404,2.528)
LB on $\mathrm{Med}(\mathcal{Y})$	0.501	(0.441,0.562)
UB on $\mathrm{Med}(\mathcal{Y})$	1.825	(1.756,1.897)
UB on $\mathrm{Med}(\mathcal{Y})$, Positive Selection[d]	1.129	(1.086,1.179)
LB on $\mathrm{Med}(\mathcal{Y})$, Negative Selection	2.367	(2.289,2.444)
Reduced Form on Z[e]	2.468	(2.404,2.528)

Data generated from 1,000 replications the model in (A.1) and (A.3) with 10,000 observations each. The parametric selection estimator of $\mathrm{E}(\mathcal{Y})$ corresponds to the constant of a regression that includes the mills ratio, generated from a first stage probit of E on Z, as a control. UB and LB abbreviate upper and lower bound. The construction of the bounds are detailed in Appendix A.2.3. The reduced form on Z entry in Panel D sums of the constant and coefficient in a regression where Z enters as the only explanatory variable. [a,b,c]: mean squared error with respect to $\mathrm{E}(\mathcal{Y})$ are 0.981 (0.818,1.158), 0.100 (0.079,0.122), and 0.760 (0.661,0.870), respectively. [d]: mean squared error with respect to $\mathrm{Med}(\mathcal{Y})$ is 0.112 (0.086,0.141). [e]: mean squared error with respect to $\mathrm{E}(\mathcal{Y}|E_1=1,E_0=1)$ is 0.001 (0.000,0.003).

Figure A.I: Simulated Data, Distribution of Wages

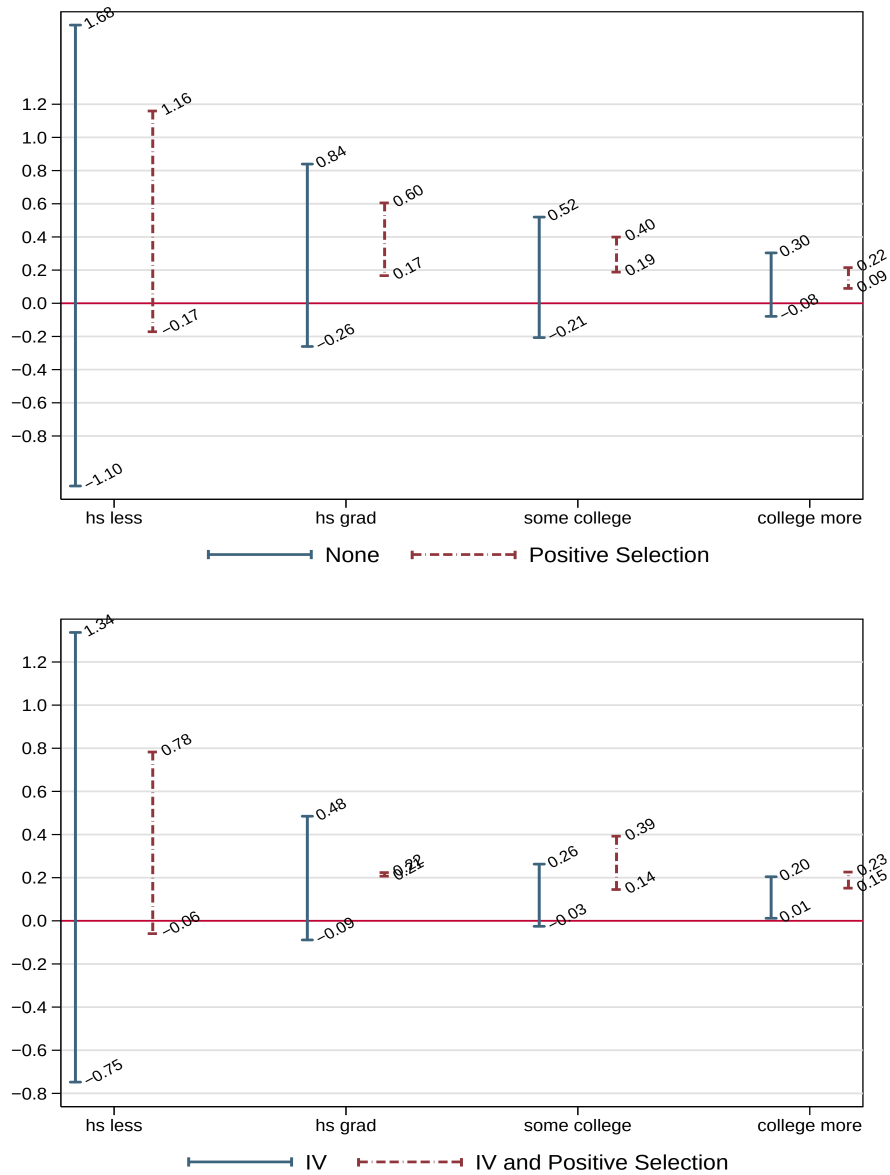
All Wages | Wages of Employed

All Wages, type N | All Wages, type P



Data: Distribution of wages in one dataset with 10,000 observations generated by the model in equations (A.1) to (A.3).

Figure A.II: Bounds on Median Wage Gap, US



Data: March CPS, ADF, 1976-2005. Sample delimitation is as in notes to Table 1, with further details in Appendix A.3. The figure plots changes in the gender median wage differential from 1976 to 2005, as specified in (A.11) and (A.12), stratified by education groups. "None" denotes neither the use of an exclusion restriction nor the use of a selection assumption when estimating the bounds. "Positive Selection" denotes the use of a positive selection assumption for all periods and all genders. "IV" denotes the use of an exclusion restriction, a binary indicator for the presence of a child younger than six, in the bounds for the median wages of women. "IV and Positive Selection" combines the exclusion restriction for women with a positive selection assumption for both men and women. The bounds under "IV and Positive Selection" for individuals with some college or more cross.