The Gender Gap Between Earnings Distributions

Esfandiar Maasoumi
Emory University

Le Wang
University of Oklahoma, & IZA

Abstract

We examine the whole distribution of the gender gap in the United States for several decades, including the most recent recession. Heterogeneity of outcomes and selection into the labor force are highlighted. Perception of any distributed gap is sensitive to its definition. We provide a brief synthesis of the decision-theoretic basis for measures of the gap and some closely related concepts in the inequality literature. Since cardinal definitions of any gap are subjective, robustness of the findings is examined by tests of stochastic dominance. A quantile-Copula approach is adopted to account for labor force participation and to recover the whole wage distributions for the male and female populations, including those who do not work full-time. Our approach robustifies some prior findings, modifies and changes others.

The magnitude and evolution of the gender gap varies by different measures. Compared to mean or median gap, our preferred (Generalized Entropy) measures indicate a generally larger convergence until early 90s, and a more pronounced flattening since then. This evolution corresponds to a decline in labor force participation (LFP) for males, and an increasing one for women which may have peaked. Our results indicate that, once selection is accounted for, the gap is no longer trending down, uniformly, suggesting a slower convergence and even a recent reversal in the trend over portions of the distribution between mid-1990s and the most recent recession. During the most recent recession, there was a marked decline in the gap among low-skilled workers, which is likely due to a relative deterioration in wages of the low-skilled males. LFP also varies by education and race. Convergence is much smaller amongst the least educated or black women, especially during the recent years. Also, non-market values, especially for those who choose not to work, are estimated, providing a new measure of the gap in the market *and* non-market distributions between men and women. Failure to account for non-market values can both exaggerate the gender gap and mask the deteriorating situation for some women. Using the estimated wage function and distributions, we further assess and challenge a variety of assumptions, hypotheses, and findings in the literature on the gender gap and inequality. These assessments explain our choices.

—

† We specially thank Professor James Heckman (the Editor) for very generous and extensive comments and guidance. We are grateful to six knowledgeable referees for their constructive comments and suggestions. We thank Stephane Bonhomme for inspiring discussions of the findings. Part of the project was undertaken during Le Wang's visit at Woman and Public Policy program at Harvard Kennedy School, and he would like to thank them for their support. We thank Sol Polachek and seminar participants including at Aarhus, InfoMetrics Institute, Georgia State, Stanford, Boston College, U Washington, UCR, USC, Bentley University, Union College, Emory University, Jinan University, Kansas State, Alabama, New Hampshire and Southern Economic Association for their helpful comments.

Contents

1. Introduction

The *distribution* of the gender gap in the US has evolved over several decades. There is heterogeneity in outcomes, and labor force participation decision has a major impact on inferences. Mindfull of these two issues, we provide a more robust perspective on the definition of the gap and its trend.

Researchers have noted that the gender gap has decreased over time, especially in the 1980s and early 1990s. This trend has slowed down since the mid-1990s. A generally encouraging trend is noted: women are catching up with men (Blau and Kahn (1997), Blau and Kahn (2006), and Goldin (2014)). There is less agreement on the magnitude and the movement of the gap over the business cycle. We offer robustness analysis based on flexible entropy summary measures and tests for uniform ranking by stochastic dominance. Full time employed men's wage distribution is preferred to the corresponding for women, even though the gap evolves differently over several decades. Efforts are also directed at decomposition/counterfactual analysis which date back to Oaxaca (1973) and Oaxaca and Ransom (1994). Recent literature has extended such decomposition methods, including to nonparametric settings (e.g., Frolich 2007; Mora 2008; Nopo 2008) and the distributional level (e.g., DiNardo et al. 1996; Blau and Kahn 1994; Albrecht et al. 2009; Fortin et al. 2011). This type of analysis aspires to decompose the gender gap into human characteristics (composition effects), and market structure effects.

The narrowing of the gender gap and its evolution are, however, sensitive to the function of the wage distribution that defines the gender gap. The magnitude and evolution of the gap can differ across the wage distribution (e.g., Albrecht et al., 2003, Blau and Kahn, 2016). Given this heterogeneity, what is a desirable integrative/scalar measure of the *overall* gender gap? We contribute to this literature by first providing a brief synthesis of the decision-theoretic basis of various measures of distance between whole distributions. This helps to integrate the inequality literature with the one on the gender gap, and empirical labor generally. The subjectivity of all cardinal measures leads to a search for *uniform* rankings by statistical tests for stochastic dominance. The latter has been widely used to analyze inequality, poverty and financial outcomes, but not to analyze the gender gap.

Other subjective functions of thr gap between wage distributions weight and compare different percentiles differently than at the mean and median. We introduce Generalized entropy measures that include a Kullback-Leibler-Theil measure and the Hellinger measure. We show that these are relative summary measures. Our entropy measures provide a nuanced picture of the evolution of the gap, compared to the mean or median. For example, Generalized Entropy measures indicate a generally larger convergence until early 1990s, and a more pronounced flattening since then, for full time workers.

These findings do not hold up once selection is accounted for. Our second contribution is to assess the impact of sample selection at the distribution level. Labor force participation (LFP) rates for males have continued to decline for decades, and those for females increased, then peaked, and has decreased slightly in recent years. To the extent that non-working men and women systematically differ from working men and women, measures of the gap would be biased. For example, if there is positive selection by women over time (high-earning women enter the labor market, and low-earning ones leave), we may observe convergence in the gender gap even though there may not be any wage adjustments. This key insight dates back to Heckman (1974), and attention has been paid to it in many studies of women's labor market outcomes. However, in the gender gap literature, we note only few attempts focused on the gap at the mean or median (e.g., Blau and Kahn, 2006; Olivetti and Petrongolo, 2008; Mulligan and Rubinstein, 2008).

We adopt a new quantile-Copula approach to the joint determination of wages and participation decision for both men and women. This approach is due to Arellano and Bonhomme (forthcoming) and can address the selection issue beyond the mean and median. It allows flexibility in specification of the marginal processes and distributions for wage outcomes and participation equations. It also enables us to relax the rank invariance assumption commonly made in the Quantile Treatment Effect (QTE) literature, and to recover the gender gap between the distributions of wage offers for the entire male and female populations. This accounts for "value of time" including those who do not work.

Participation decision is consequential and gender gap evolves differently once corrected for selection. Convergence is slower, with a recent reversal in the trend in parts of the wage distribution between mid-1990s and the most recent recession. During the most recent recession there was a marked decline in the gap among low-skilled workers. This is likely due to a relative deterioration in the wages of low-skilled males rather than relative rises in wages of low skilled women. This three-phase trend is masked if selection is unaccounted for.

Once selection is accounted for, uniform ranking of wage distributions between men and women is less likely, and we do not find a weakly "uniform" narrowing of the gap at all quantiles. Uniform ranking tests is reported for a range of observed and counterfactual distributions.

Labor force participation varies by education and race. For example, we find that the relative economic position of less educated women lacked progress or even deteriorated during more recent years, and the existing studies may have understated this because many low-wage earners among less educated women exit the labor force. Similar results hold for black women. Specifically, the wage gap for black women has narrowed less compared to both Hispanics and whites, although the gap within minority groups (blacks and Hispanics) is generally smaller than the gap amongst whites.

Our third contribution is to consider value of time. Comparing distributions of wage *offers* is informative, but for women who do not work, wage offers do not reveal "value of time". Some individuals derive value from not working, and this is captured by their reservation wages. To examine this, we provide an additional concept of the gender gap which replaces women's wage distribution with a mixed distribution of individuals' wages conditional on employment, and individuals' *reservation wages* conditional on non-employment. The quantile-copula approach provides a useful structure to recover the reservation wages and its distribution using the potential wage offers and the selection mechanism. This provides new results that have not been previously discussed and explored. For example, taking into account value of time, women's relative well-being, especially among those in the upper tail, may have worsened over time.

Our approach is premised on a number of methodological choices. Using our estimated wage functions and distributions, we are able to assess and challenge a variety of alternative assumptions, hypotheses, and findings in the existing literature. First, with estimated selection parameters, we are able to trace out the evolution of selection mechanisms into labor market for both men and women and relate it to the long-run trend in the labor force participation in the U.S., especially during the most recent recession. The evidence indicates that there has been a significant difference in selection between men and women, and that there has been a fundamental change in the selection pattern for women over time, moving from negative to positive selection. In the presence of both positive and negative selection, 1) selection is not systematically related to the employment rates among women and, 2) plays a limited role in explaining the observed relationship between employment- and wage gaps between genders. Second, we test a popular dominance (monotonicity) assumption that is often imposed to obtain bounds on the wage distributions in the presence of sample selection, as in Blundell et al. (2007), and show that it is rejected for the US data for the majority of the samples. Third, with the distributions of potential wages that we recover, we examine the robustness of various inequality measures to the presence of sample selection. While we confirm that inequality among both men and women has been generally increasing over time, whether or not selection is accounted for, we also challenge the conventional wisdom that the increased overall inequality amongst men is only attributed to the increasing trend in the upper tail, but not the lower tail.

Finally, we derive relevant counterfactual distributions that can shed light on potential explanations of the gender gap. Our results suggest that failure to account for selection may underestimate the importance of "skills" but overestimate the importance of market structure in explaining the gender gap.

The rest of the paper is organized as follows. Section 2 lays out the methodology and decision theoretic bases of definitions of the gap. Section 3 discusses the data and presents the baseline results. These results facilitate comparison to the literature, and highlight heterogeneity of the gap and its evolution over time, as well as the sensitivity to different definitions of the gap. Section 4 provides a model of selection and its implementation and examines selection results to be compared with baseline findings. In section 5., we assess the gap for different subgroups by education and race, controlling for labor force participation decision. In Section 6, we propose a new gender gap measure accounting for non-market (time) value for men and women who do not work full time. To justify our methodological choices, in Section 7, we assess a variety of assumptions, hypotheses, and findings in the existing studies of the gender gap and inequality. Section 8 summarizes some of the main findings and contributions and applications.

2. Definition of the Gap

Despite significant heterogeneity in wages, the gap is often reported between average wages and medians.There is considerable improvement when select percentiles are also reported.[1] Averages give equal weight to all percentiles, median and any single quantile imply zero weights to other quantiles. Unintended impressions of uniform ranking can flow from this practice. All summary/scalar measures assign subjective weights to different quantiles and embody inter-personal comparisons.[2]

[1]For example, in Blau and Kahn (2016), the most recent, comprehensive survey of the gender gap, three select percentiles are examined.

[2]We refer readers to survey articles on this topic and references in other fields for more detailed discussions of these alternative measures and their welfare and decision-theoretic bases (e.g., Atkinson, 1970; Aaberge et al., 2013; Maasoumi, 1993; Maasoumi, 1998; Kolm, 1969).

2.1. Decision-Theoretic Basis of Measures of Gap

Let y^f and y^m denote (log) wages of females and males, with CDF (density) denoted by F_f (f_f) and F_m (f_m), respectively. Let $F_f(y^f_\tau) = \tau$, and $F_m(y^m_\tau) = \tau$ define the τ-th quantile. The gap at a τ^{th} quantile is $y^m_\tau - y^f_\tau$, including the median ($\tau = \frac{1}{2}$). Mean gap is $\mathbb{E}[y^m] - \mathbb{E}[y^f] = \int_0^1 \left[y^m_\tau - y^f_\tau \right] d\tau$. Gap at any quantile and at the mean are examples of (linear) weighted functions of quantile gaps. Linear summary functions of quantiles are subjective. Other functions would reflect different subjective weights and/or interpersonal evaluations, reflecting degrees of aversion to inequality/dispersion and/or risk. There is a large literarture on risk and inequality measures. To be brief, consider the following general class of Evaluation Functions (EFs):

$$EF_{\gamma,\epsilon} = \int_0^1 R(\tau,\gamma) U_\epsilon(y_\tau) d\tau \tag{1}$$

where $R(\tau,\gamma) = \gamma(1-\tau)^{\gamma-1}$, and $U(\cdot)$ is a concave function of wages. γ is an aversion parameter defined further below. This class of functions is general enough to include both the Atkinson and S-Gini families of inequality measures which satisfy desirable properties such as the Pigou-Dalton transfer and population properties. If only relative (scale/mean) independent measures are to be considered, the function $U(\cdot)$ must be of the following (homothetic) form:

$$U_\epsilon(y_\tau) = \begin{cases} \frac{y_\tau^{1-\epsilon}}{1-\epsilon} & \text{if } \epsilon \neq 1 \\ \log(y_\tau) & \text{if } \epsilon = 1 \end{cases} \tag{2}$$

See Arrar and Duclos (2003) or Pratt (1964). Note that the wage quantile y_τ itself is the most popular example of the utility function $U(\cdot)$ at $\epsilon = 0$. An important example of a money metric Evaluation Function is the Equally Distributed Equivalent (EDF) wage,

$$\begin{aligned} EDF_{\gamma,\epsilon} &= U^{-1}(EF_{\gamma,\epsilon}) && (3) \\ &= \mu_y \left(1 - I_{\gamma,\epsilon}(y)\right) && (4) \end{aligned}$$

where μ_y is the mean and $I_{\gamma,\epsilon}(\cdot)$ is any relative inequality measure. There are many inequality measures! A strongly justified monotonic transformation of the Atkinson family of inequality indices is the Generalized Entropy (GE) family as follows,

$$I_\gamma = \frac{1}{\gamma(\gamma+1)} \int_0^1 (\frac{y_\tau}{\mu_y}) \left[(\frac{y_\tau}{\mu_y})^\gamma - 1 \right] d\tau, \gamma \in R \tag{5}$$

where γ is the degree of aversion to relative inequality/risk; and for $\gamma < 2$, the higher its absolute value the greater is the sensitivity to transfers in the lower tail of the distribution. Such measures are generally a complicated weighting function of quantiles and the underlying distribution.

A definition of the gender gap that subsumes means, medians, and more complex functions of the quantiles is the difference of respective EFs:

$$\text{Gender Gap } = EF_{\gamma,\epsilon}(y^m) - EF_{\gamma,\epsilon}(y^f) \tag{6}$$

Its relation to inequality and entropy measures is immediate. A relative measure of the gap (or when mean wages are equal) is merely the difference of entropy inequality measures. Each measure $I(x)$ is itself the divergence between the entropy of the wage distribution and (max) entropy of the uniform distribution (equality). A relative measure of the gap is thus the entropy *divergence* between the wage distributions (the uniform distributions cancel out). These are well known (generalized) Kullback-Leibler type measures of "information gain" which are generally asymmetric and also fail the trainangle inequality (are not metric). They can be symmetrized and include the normalized Hellinger measure which is a metric. For $\gamma = -1$:

$$\text{Theil-KL} \;=\frac{1}{2}\cdot(\frac{1}{2}\cdot\int[\log(\frac{f_f}{f_m})\cdot f_f+\log(\frac{f_m}{f_f})\cdot f_m]dy) \tag{7}$$

For and $\gamma = \frac{-1}{2}$, a normalization of the Bhattacharya-Matusita-Hellinger measures is given by:[3,4]

$$\begin{aligned} S_\rho &= \frac{1}{2}\int_{-\infty}^{\infty}\left(f_m^{1/2}-f_f^{1/2}\right)^2 dx \\ &= \frac{1}{2}\int\left[1-\frac{f_m^{1/2}}{f_f^{1/2}}\right]^2 dF_m \end{aligned} \tag{8}$$

KL-Theil measure is more "inequality averse" than Gini and the Hellinger.[5].

The above are relative measures, and money-metric evaluations can be derived from Equation (6), and other monotonic transformations. A representation of $EF_{\gamma,\epsilon}$ (using integration by parts) reveals a very useful relation to Second Order Stochastic dominance (SSD) which will be tested for and reported:

$$EF_{\gamma,\epsilon} = \int_0^1 \gamma(\gamma-1)(1-\tau)^{\gamma-2}GL_U(\tau)d\tau \tag{9}$$

where $GL_U(\tau) = \int_0^\tau U(y_u)du$ is the Generalized Lorenze function of $U(\cdot)$. When $U(\cdot) = y_\tau$, ranking by the corresponding GL is the same as ranking by Second Order Stochastic Dominance.

To summarize, the S_ρ and symmetrized Theil measures of the gap are scale invariant and are not affine functions of quantiles.[6] Affine functions imply infinite substitutability between quantiles.

[3]Note that in addition to being metric, this measure also satisfies many desirable properties: 1. is well defined for both continuous and discrete variables; 2. is *normalized* to $[0, 1]$; is well defined and applicable when X is multidimensional; 3. is *invariant* under continuous and strictly increasing transformations, such as logarithmic. This feature can be particularly useful in this context (see, e.g., footnote 16).

[4]Following Granger et al. (2004) and Maasoumi and Racine (2002), we consider a nonparametric kernel-based implementation of (8) (The computer code **-srho-** written by the authors in Stata is also available upon request). In our study, we use Gaussian kernels and the "normal reference rule-of-thumb" bandwidth ($= 1.06\min(\sigma_d, \frac{IQR^d}{1.349}) * n^{-1/5}$, where $\sigma_d, d = m, f$ is the sample standard deviation of $\{\ln(w_i^d)\}_{i=1}^{N_d}$; IRQ^d is the interquartile range of the sample d.). Integrals are numerically approximated by the integrals of the fitted cubic splines of the data, which "give superior results for most smooth functions" (StataCorp, 2009). We employ bootstrap re-sampling procedure based on 299 replications.

[5]Gini is, effectively, "blind" to the tails! Evaluation of policies aimed at the tails (e.g., anti poverty) will look in vain for changes in the Gini!

[6]Note the underlying preference itself does not change, but because the functions are NOT affined functions of quantiles, the (nonlinear) weights on quantiles depend on the distribution and hence vary. The changes are due to changes in the distributions, but not the differences in the preference function.

2.2. Uniform Ordering: Stochastic Dominance Tests

Subjective indices of the gap provide "complete" (cardinal) rankings. When distributions cross (especially at lower tails)[7], different measures will differ in their rankings. It is useful to test whether distributions can be uniformly ranked over large classes of (evaluation) functions to a statistical degree of confidence. Absent any uniform dominance relations, integrative entropy measures like S_ρ and other quantiles need to be examined.

Let U_1 denote the class of all *increasing* von Neumann-Morgenstern type utility functions u that are increasing in wages (i.e. $u' \geq 0$), and U_2 the class of utility functions in U_1 such that $u'' \leq 0$ (i.e. concave). Concavity implies an aversion to inequality:

First Order Dominance:

Male wages y^m First Order Stochastically Dominate (FSD) Female wages y^f *if and only if*

1. $EF(y^m) \geq EF(y^f))$ for all $EF \in U_1$ with strict inequality for some EF;
2. Or, $F_m(y) \leq F_f(y)$ for all y with strict inequality for some y.
3. Or, $y^m_\tau \geq y^f_\tau$ for all points on the support.

Second Order Dominance:

Male wages (y^m) Second Order Stochastically Dominates Female wages (y^f) (denoted y^m SSD y^f) *if and only if*

1. $EF(y^m) \geq EF(y^f)$ for all $EF \in U_2$ with strict inequality for some EF;
2. Or, $\int_{-\infty}^{y} F_m(t)dt \leq \int_{-\infty}^{y} F_f(t)dt$ for all x with strict inequality for some x.
3. Or, $\int_{-\infty}^{y} y^m_\tau dt \geq \int_{-\infty}^{y} y^f_\tau dt$ for all points on the support.

x^m FSD x^f implies that the mean male wage is greater than the female mean wage. FSD implies SSD. Higher order SD rankings are based on narrower classes of preferences. [8]

[7] as they do for wages in some years after correcting for selection.

[8] We employ SD tests based on a generalized Kolmogorov-Smirnov test discussed in Linton et al. (2005). The tests for FSD and SSD are based on the following functionals:

$$d = \sqrt{\frac{N_1 N_2}{N_1 + N_2}} \min \sup [F_m(y) - F_f(y)] \tag{10}$$

$$s = \sqrt{\frac{N_1 N_2}{N_1 + N_2}} \min \sup \int_{-\infty}^{y} [F_m(t) - F_f(t)]dt \tag{11}$$

N_1 and N_2 are respective sample sizes. Test statistics are based on the sample counterparts of d and s, employing empirical CDFs. We use bootstrap implementation of the tests for iid samples. We estimates the probability of the tests falling in any desired interval, as well as p-values. If the probability of d lying in the non-positive interval (i.e. $Pr[d \leq 0]$ is large, say .90 or higher, and $\widehat{d} \leq 0$, we can infer FSD to a high degree of statistical confidence. Maasoumi (2001) surveys the related tests and techniques.

3. Baseline Results

3.1. Data

We examine the period 1976-2013, March Current Population Survey (CPS). We use log of hourly wages, measured by an individual's wage and salary income for the previous year divided by the number of weeks worked and hours worked per week. [9]

Our sample includes individuals aged between 18 and 64 who work only for wages and salary, and do not live in group quarters. In our baseline results which ignore selection, our sample includes only those who worked for more than 20 weeks (inclusive), and more than 35 hours per week in the previous year (e.g., Mulligan and Rubinstein 2008). Information about sample size is provided in the supplemental material (Table C.1). [10]

3.2. Baseline Estimates of the Gender Gap 1976 - 2013 for the Employed Sample

Table (1) reports a number of measures of the gender gap. Column (1) displays our metric entropy measure of the gender gap. S_ρ is normalized, taking values in $[0,1]$, and to facilitate the presentation, it is multiplied by 100. Columns (2) and (7) display the difference of log earnings at select percentiles between men and women (including mean and median). All are statistically significant. The standard errors based on 299 resamples are reported in the supplemental material.

The gap is substantial and positive. The implied *Dollar* differentials vary at different quantiles and at the mean. In 1976, the average gender gap at the 10^{th} percentile is about 31 percentage points, about 50 percentage points at the 90^{th} percentile, and varies between 43 and 46 percentage points at other percentiles. This is an example of an heterogenous but uniformly positive gap. The arithmatic "average" uses *equal weights* at all earnings and is sensitive to outliers. Geometric mean is known to be better but is never reported, and genralized geometric mean requires subjective parameters.

The gap decreased over the past four decades by all measures, but not monotonically. The timing of temporal deviations from the long-run trend varies across different measures. The conventional measures of the gap do not generally move in the same direction, except in few years (1980, 1989, 1994, and 2002). For example, in 1977, the gap at the median and 70^{th} percentile increased, while it decreased at other parts.[12]

The S_ρ is statistically significantly different from zero in all cases. It is decreasing in 1977, consistent with the decrease at all parts but the median and 70^{th} percentile; it increased in 1999, consistent with the increase only at the 10th and 75th percentiles. The entropy measure is able to account for *not only* increasing earnings *but also* the greater dispersion (inequality increasing) accompanying it. In 2009 (the

[9] Wages are adjusted for inflation based on the 1999 CPI adjustment factors. These are available at `https://cps.ipums.org/cps/cpi99.shtml`. Following the literature (e.g., Mulligan and Rubinstein 2008; Lemieux 2006), we exclude extremely low values of wages (less than one unit of the log wages). It has been shown that *inclusion* of imputed wages in wage studies is "problematic" (Hirsch and Schumacher 2004; Bollinger and Hirsch 2006). Mulligan and Rubinstein (2008) and Lemieux (2006)) exclude these imputed observations. Such corrections, though simple, are considered to "largely eliminate the first-order distortions resulting from imperfect matching" (Bollinger and Hirsch 2013).

[10] We use person-level weight (WTSUPP) variable throughout our analysis (which "should be used in analyses of individual-level CPS supplement data").[11] We also repeat all our analysis using NBER Outgoing Rotation Group data to assess the robustness of our results (See footnote 18), as suggested by a referee.

[12] All these results are also conveniently depicted by patterns of changes in different measures in Table (C.2). The cells with "I" highlighted in green are the years when the gap increased, the cells with "D" are the years when it decreased.

Table 1: Measures of The Gender Gap (fulltime employed)

Year	S_ρ (1)	Theil (2)	Mean (3)	10th (4)	25th (5)	50th (6)	75th (7)	90th (8)
1976	10.566	22.230	0.432	0.311	0.427	0.461	0.461	0.486
1977	10.110	21.327	0.419	0.297	0.397	0.470	0.465	0.470
1978	10.247	21.619	0.425	0.272	0.392	0.465	0.472	0.465
1979	10.127	21.173	0.421	0.262	0.386	0.468	0.487	0.474
1980	9.857	20.437	0.408	0.252	0.382	0.446	0.475	0.448
1981	9.221	19.423	0.395	0.263	0.338	0.447	0.477	0.440
1982	8.823	18.547	0.394	0.254	0.345	0.448	0.467	0.443
1983	7.645	15.992	0.376	0.240	0.323	0.427	0.439	0.451
1984	6.747	14.115	0.356	0.247	0.298	0.412	0.427	0.405
1985	6.258	12.976	0.343	0.208	0.288	0.357	0.401	0.414
1986	5.607	11.550	0.327	0.189	0.297	0.366	0.387	0.393
1987	5.023	10.292	0.322	0.214	0.288	0.376	0.365	0.378
1988	4.526	9.271	0.305	0.222	0.264	0.329	0.353	0.354
1989	4.215	8.729	0.298	0.210	0.244	0.323	0.324	0.352
1990	3.645	7.418	0.283	0.183	0.265	0.303	0.336	0.324
1991	3.109	6.355	0.254	0.143	0.208	0.273	0.307	0.330
1992	2.784	5.640	0.241	0.107	0.211	0.251	0.285	0.299
1993	2.518	5.116	0.228	0.145	0.194	0.262	0.274	0.285
1994	2.153	4.367	0.215	0.142	0.175	0.235	0.262	0.281
1995	2.048	4.157	0.221	0.143	0.197	0.239	0.283	0.276
1996	2.165	4.471	0.231	0.148	0.186	0.240	0.248	0.264
1997	2.071	4.209	0.227	0.158	0.210	0.241	0.262	0.259
1998	2.082	4.204	0.228	0.131	0.214	0.261	0.250	0.274
1999	2.236	4.616	0.235	0.154	0.192	0.231	0.251	0.262
2000	2.059	4.175	0.232	0.182	0.202	0.247	0.285	0.284
2001	1.913	3.916	0.226	0.143	0.192	0.227	0.261	0.300
2002	1.781	3.676	0.216	0.117	0.185	0.205	0.236	0.278
2003	1.684	3.434	0.208	0.125	0.147	0.203	0.238	0.265
2004	1.353	2.687	0.188	0.130	0.153	0.164	0.230	0.288
2005	1.393	2.886	0.190	0.105	0.182	0.180	0.219	0.255
2006	1.321	2.631	0.191	0.113	0.145	0.182	0.204	0.243
2007	1.102	2.179	0.180	0.125	0.153	0.183	0.206	0.227
2008	1.149	2.284	0.172	0.118	0.103	0.192	0.214	0.229
2009	1.275	2.568	0.186	0.136	0.148	0.163	0.212	0.270
2010	1.275	2.603	0.187	0.131	0.134	0.173	0.220	0.254
2011	1.118	2.275	0.175	0.111	0.140	0.173	0.211	0.236
2012	1.129	2.246	0.178	0.111	0.139	0.170	0.215	0.266
2013	0.974	1.994	0.167	0.108	0.113	0.172	0.198	0.223

[1] Note that S_ρ and Theil are multiplied by 100.

great recession), the average gender gap increased, while the median gap decreased! This conflicting result could lead to different conclusions about the cyclicality of the overall gap. The entropy measure suggests that the *overall* gap increased (possibly due to worsened economic conditions), in agreement with the mean. Interestingly, the gap at the 10^{th} percentile fluctuates more than other quantiles; the gender gap at the 90^{th} percentile exhibits a consistently declining trend with some fluctuation. The equivalent KL-Theil measure of the gap generally agrees with the metric entropy measure (but is generally larger due to its greater aversion to inequality).

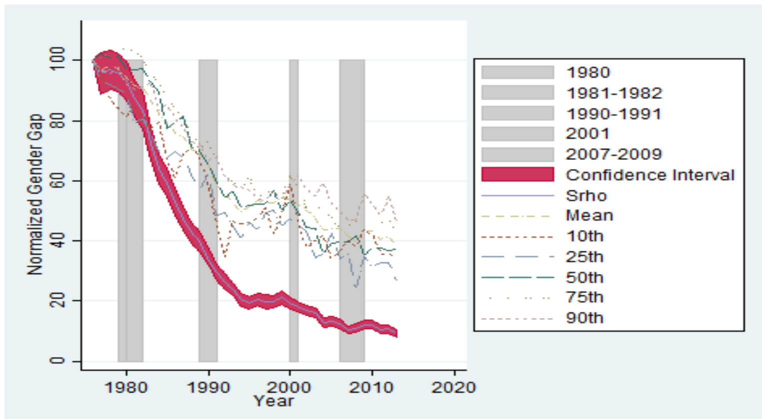


Figure 1: The Trend of Gender Gap (Shaded Areas correspond to the recession periods announced by NBER)

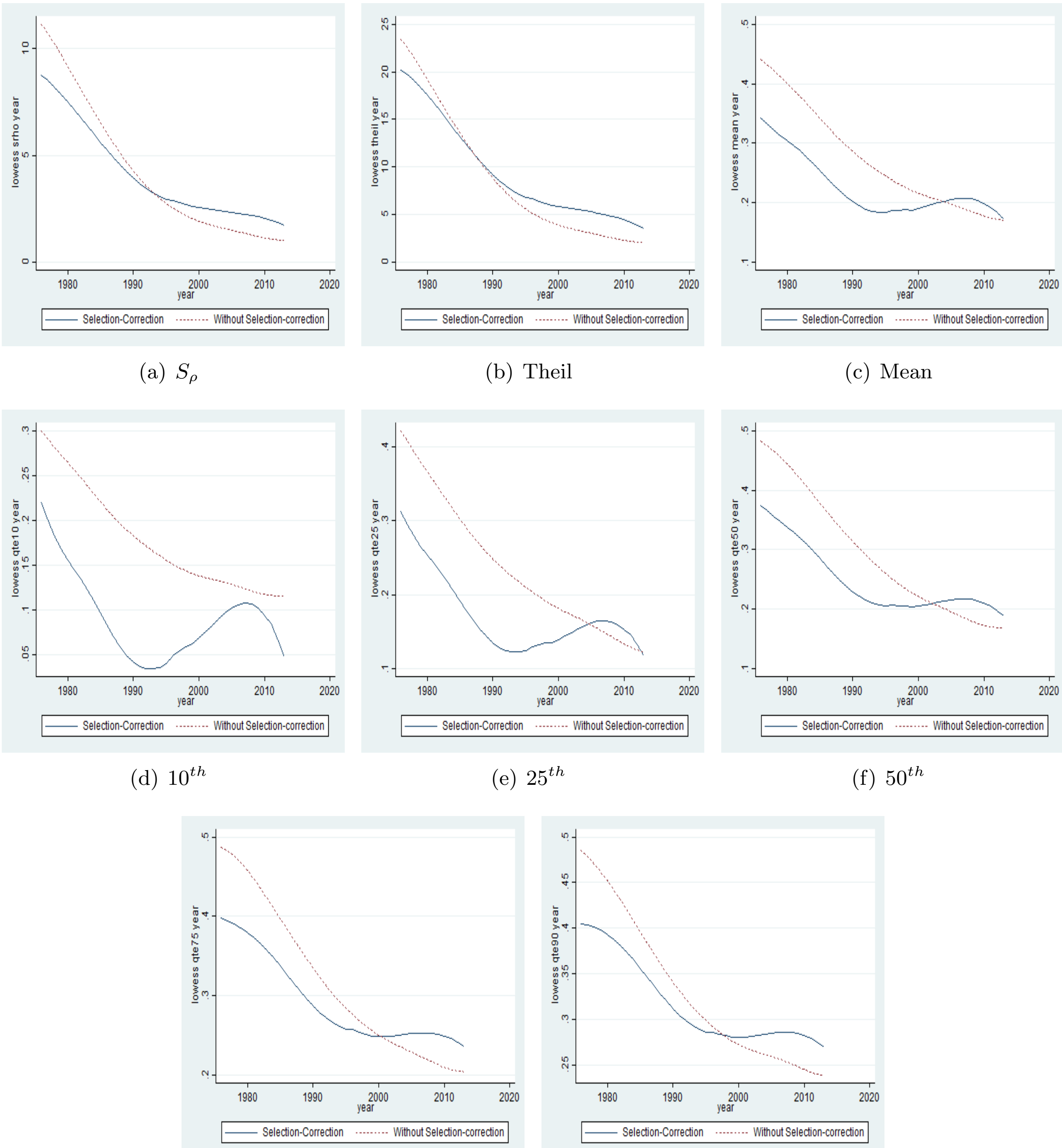
The normalized gap measures are depicted in Figure (1), with confidence intervals for S_ρ in (red) shaded areas. [13] Conventional measures of the gap correspond differently to business cycles, with varying directions. The entropy gap is relatively robust to recessions (except in 2001) as it accounts for inequality and integrates different movements in the quantile gaps.

3.2.1. The Long-Run Trend Implied by Various Measures

In Figure (2) we present smoothed trend lines for each measure using lowess (i.e., a locally weighted regression of time). The movements implied by conventional measures are in broad agreement with the literature (e.g., Blau and Kahn, 2006). The gender gap fell rapidly until the early 1990s, continued a general downward trend at a much slower rate until the most recent recession, and remained relatively stagnant with somewhat modest declines (at best) afterward. Convergence in the lower tail is somewhat larger than in the upper tail. Full-time employed lower wage female workers caught up more quickly with their male counterparts than high-wage women. This is consistent with Blau and Kahn (2016).

The long-run trend implied by our entropy measures generally agree with conventional measures. However they suggest much larger declines and rates of decline: the gender gap dropped *precipitously* before 1990s, but the trend of "convergence" slowed down since 1990s. Traditional measures understate the decline in *overall* gender gap over time. The entropy measures decline by about 90 percent, other measures by about 50 percent, over the entire period. Table (2) further quantifies the differences in

[13] We normalize the measures in Table (1) by setting the values in 1976 to 100, and generate normalized values based on original growth rates. Recessions dates are those announced by the National Bureau of Economic Research and shaded in Figure (1).



(a) S_ρ

(b) Theil

(c) Mean

(d) 10^{th}

(e) 25^{th}

(f) 50^{th}

(g) 75^{th}

(h) 90^{th}

Figure 2: Comparison of Smoothed Trend of The Gender Gap with and without Selection Correction (excluding 2010)

Table 2: Implied Long-Run Annual Changes in the Gender Gap

Period	S_ρ (1)	Theil (2)	Mean (3)	10th (4)	25th (5)	50th (6)	75th (7)	90th (8)
1973-2013	-0.061	-0.062	-0.025	-0.025	-0.032	-0.027	-0.023	-0.019
1973-1994	-0.067	-0.068	-0.029	-0.032	-0.034	-0.03	-0.026	-0.024
1994-2013	-0.051	-0.051	-0.019	-0.016	-0.028	-0.023	-0.017	-0.012

[1] These values are long-run compound annual change rates implied by the initial and the last smoothed values of each period.

the implied long-run trends, reporting implied compound annual convergence rates over the entire period [14]. [15] [16].

3.2.2. Stochastic Dominance Rankings

We examine empirical distributions of earnings for men and women without regard for age, education, or other relevant factors. Baseline SD rankings are similar across years. We exemplify these results with the most recent year in Table (3). All results and the graphical comparisons of CDFs are reported in the supplemental material. The column labeled *Observed Ranking* indicates if distributions may be ranked in either the first or second order; the columns labeled $Pr[d \leq 0]$ and $Pr[s \leq 0]$ report the p-values based on the simple bootstrap technique. If we observe FSD (SSD) *and* $Pr[d \leq 0]$ ($Pr[s \leq 0]$) is large, say 0.90 or higher, we may infer FSD to a degree of statistical confidence.[17]

These findings provide strong robustness for views on full-time employed workers: anyone with an Evaluation Function in the class U_1 (merely increasing in earnings) would prefer the male distribution to the female distribution. Since FSD implies higher order rankings, inequality aversion does not reverse the rankings. *Despite a narrowing gap, women do not perform better than men across the whole distribution.* These rankings are not sustained once selection is accounted for. They also change

[14] Specifically, the implied annual change is calculated from the equation: Last value $= (1+r)^T \cdot$ Initial Value, where T is the number of years during this time period.

[15] The entropy measures of the overall gap indicate an annual percentage change of about 6% for 1976-2013, while the long-run annual convergence rates implied by the conventional measures vary between 2% to 2.7%. Within sub-periods, the annual percentage changes in entropy measures were 7% before 1994, and about 5% afterwards. Welfare reforms were enacted by many states in the mid-1990s and by Congress in 1994 (Waldfogel and Mayer, 2000). Moreover, there was a new wave of skill-biased technological progress during the 1990s and "a marked acceleration in technology" in the period 1995-1999 (Basu et al., 2001). One might expect both welfare reform and technological progress to accelerate convergence. Our results indicate otherwise. As will be shown below, accounting for selection for both men and women reinforces this observation.

[16] **Log vs Level**: Entropy gap is invariant to logarithmic and monotonic transformation of wages. Other metrics reported here are not and depend on whether actual wages, or their logarithm, or some other transformation is used. As suggested by a referee, we plot the (normalized) conventional measures of the gap using both the levels and logs of wages in Figure (C.1) in the supplemental material. While the implied pattern by the gap at $10^{th}, 25^{th}$ and 50^{th} percentiles is relatively similar between levels and logs, the rate of decline using levels is slightly larger than that implied by logs; this is particularly true for the mean gap. For the data here the discrepancies are relatively modest. However, it is conceivable that in other contexts such differences may be large.

[17] The CDFs (Appendix C, Supplemental Material) for men lie predominantly to the right of the ones for women, indicating higher level of earnings for men at all sample quantiles. Moreover, the earnings distributions for men and women move closer over time, consistent with our measures of the gender gap.

Table 3: An Example of Stochastic Dominance Results (Female v.s. Male Wage Distributions)

Year	Observed Ranking	$d_{1,max}$	$d_{2,max}$	d	$Pr[d \leq 0]$	$s_{1,max}$	$s_{2,max}$	s	$Pr[s \leq 0]$
2013	FSD	14.42	-0.64	-0.64	1.00	2018.33	-0.64	-0.64	1.00

[1] We find first-order dominance in all years. The FSD ranking in 2013 is representative.

when we report counterfactual results with controls for some characteristics.[18]

4. Accounting for Selection

Female labor force participation in the US, although rising over time, remains low compared to men, and to women in other developed countries (Blau and Kahn 2013). This is also true for full-time employment among women. Based on our sample, despite a rapid increase from the beginning of our sample through 2001, women's full-time employment remains at roughly 53%. The trend has reversed after 2001, partly due to the increase in stay-at-home mothers (Cohn et al. 2014). On the other hand, male labor force participation has continued to decline over the same period (Council of Economic Advisers, 2016). Using our data, we find that men's full-time employment rate is roughly 67%.

If wages for non-working women are systematically lower than for full-time workers (positive selection), the baseline gaps under-estimate the true gender gap, while stochastic dominance rankings may remain unchanged and even strengthened. But if the wages among non-working women are higher than full-time workers (negative selection), baseline measures over-state the severity of the gender gap. The direction of the bias would be less clear once both non-working men and women are taken into account.

Blau and Kahn (2006) and Olivetti and Petrongolo (2008) recover the "true" *median* wages for women using the fact that median wages are not much affected by inclusion of imputed values that are either lower or upper bounds of the wages. These authors find that selection bias affects the

[18] **Alternative Data Source: Outgoing Rotation Group (ORG) Data.** When studying the time trends in residual wage inequality, Lemieux (2006) find that there are some differences between using March CPS and ORG data, especially among women. Lemieux (2006) notes that the difference may be due to the fact that "measurement problems in the March CPS have been magnified by the growth over time in the fraction of workers paid by the hour." We thank one referee for pointing out several issues related the CPS data discussed here and elsewhere in the paper. To assess the impact of such issue on the time trends of our gender gap measures, we repeat all of our analysis using the ORG data from NBER website.(`http://www.nber.org/data/morg.html`) We use the same sample restrictions as above and focus on full-time workers. The data contain more accurate information on hourly wages. The results are reported in Figure (C.2) and Table (C.4) in the supplemental material. We summarize the findings here. First, the normalized measures of the gender gap between March and ORG CPS data are strikingly similar. The time trends using different datasets trace out each other very closely, especially for our entropy measure. They also show that there was a rapid decline before the 1990s, slowed down afterward and remained stagnant during the most recent recession. Recall that our gender gap measure captures the whole distribution, so they should be very similar at most parts of the distributions. Second, our SD results are not sensitive to the alternative dataset. We again find that men's wage distributions dominate, in a first-order sense, women's wage distribution, indicating women perform worse than men in the labor market. The ORG data do not consistently provide the information on identification sources needed for addressing the selection issue below. We thus compare the benchmark results without addressing selection. However, the striking similarity in the implied patterns does provide some confidence in the robustness of our results using the CPS data.

observed gender gap to some extent. Blau and Kahn (2006) find that the rapid decline of the median gap in the 1980s may be overstated because of selection. Their finding was based on assumptions regarding "the position of the imputed wages with respect to the median of the wage distribution". These imputations were based on observable characteristics such as education and experience, and "selection on unobservables are assumed away" (Machado, 2012). Olivetti and Petrongolo (2008) implicitly assume a fixed selection rule, which may be invaid. By contrast, Mulligan and Rubinstein (2008) allow for selection on unobservables which is time-varying. Their focus is on the *mean* gender gap allowing Heckman type correction. They too find that selection is important in explaining the mean gap, and selection varies over time, from negative to positive. Instead of a parametric selection model, Blundell et al. (2007) employ economic theory to derive bounds on the gender wage gap and derive bounds for the gender gap at different parts of the distribution. They assume, however, a fixed, positive selection rule that working women's wages first order dominate non-working women's. Employed women have higher wage offers than non-employed women. This assumption may be too restrictive and fail to hold. We are able to test, and find that it is rejected for the US in most cases. Indeed, evidence of negative selection has also been documented in Neal (2004) and Mulligan and Rubinstein (2008).

4.1. Econometric Methods to Address Selection

We address selection at the *distributional* level and allow for time-varying selection. Our solution is based on a two-step procedure that recovers the marginal distributions of the wages from conditional quantiles, after the latter have been adjusted for selection. Recovery of marginal distribution from conditional quantiles is addressed in Machado and Mata (2005). The asymptotic statistical theory recently developed in Chernozhukov et al. (2013) may be applied. Our procedure differs from Machado and Mata (2005) and some subsequent analysis in that we take into account selection when estimating the conditional quantiles. There are a few methods in the literature to *point* identify conditional quantiles with selection. An approach proposed in Arellano and Bonhomme (forthcoming) has many advantages and is adopted here. This approach is semi-parametric and models the joint distribution of the true (or latent) quantile of the wage distribution and the participation decision, leaving a good deal of fllexibility on marginal processes for wages and selection. We do not impose the restriction in Blundell et al. (2007). We are able to assess the magnitude of selection with a parameter which captures it, as well as the change in the selection pattern over time. This further helps explain the evolution of the *observed* gender gap.

4.1.1. Preliminaries: Recovering Unconditional Distributions from Conditional Quantile Functions

Let $F_{y|x_i} \equiv Pr[y_i \leq y|x_i]$ be the conditional CDF of the wages given $x = x_i$, and $Q_\tau(y|x_i)$ the corresponding τ^{th} conditional quantile. Note that $Q_\tau(y|x_i) = F^{-1}_{y|x_i}(\tau)$, the inverse of the conditional CDF. The marginal distribution is related to conditional distribution as follows [19]

$$F_y(y) = \mathbb{E}[I[y_i \leq y]] = \mathbb{E}[\mathbb{E}[I[y_i \leq y]|x_i]] = \mathbb{E}[F_{y|x_i}] \tag{12}$$

[19]The first equality follows from the definition of unconditional CDF. And the second equality follows directly from the law of iterated expectations. The last is again the definition.

where $I[\cdot]$ is an indicator function. The conditional CDF is related to its inverse (the conditional quantile function) as follows (see, e.g., Angrist and Pischke 2009, p.282)

$$F_{y|x_i} = \int_0^1 I[F^{-1}_{y|x_i}(\tau) \leq y]d\tau = \int_0^1 I[Q_\tau(y|x_i) \leq y]d\tau \tag{13}$$

The unconditional CDF may be estimated by

$$\widehat{F}_y(y) = \frac{1}{N}\sum_{i=1}^{N}\int_0^1 I[Q_\tau(y|x_i) \leq y]d\tau] \tag{14}$$

And the corresponding unconditional quantiles can be, again, obtained by inverting the marginal CDF:[20]

$$Q_\tau(y) = \inf\{y : \widehat{F}_y(y) \geq \tau\}$$

Using these estimates, we calculate the gender gap functions as well as test for stochastic dominance. Counterfactual distributions are also defined here and derived accordingly.

4.1.2. Conditional Quantile Selection Models

In the absence of selection, a probability re-weighting approach can be used to recover marginal distributions (see, e.g., Firpo 2007). Reweighting and quantile approaches are equally valid (Chernozhukov et al. 2013). They lead to numerically identical results asymptotically. However, the reweighting approach cannot easily accommodate the selection issue. One cannot identify distributions for groups whose wages are not observed for non workers. On the other hand, the quantile-copula function approach, by adding (slightly) more structure and hence information, can address selection and enables identification. [21]

Buchinsky (1998) proposed a control function approach to extend Heckman's selection approach to quantiles. He assumed additive separability of observable and unobservables in the wage equation. It also implicitly assumed "independence between the error term and the regressors conditional on the selection probability." (Melly and Huber, 2008) Arellano and Bonhomme (forthcoming) and Arellano

[20] In practice, the pdf and CDF are obtained using the simulation method proposed in Machado and Mata (2005) and Melly (2005) for more details. First, we simulate a sample from the conditional distribution at the given covariate values, corrected for selection. Then, we integrate out these covariates to obtain a sample that is consistent with the desirable marginal distribution. Any characteristics of the distribution, including the mean, can thus be obtained based on this drawn sample, for example, the pdf with robust nonparametric kernel density estimation on the data.

[21] Parametric estimation of quantiles is due to Koenker and Bassett (1978), and nonparametric extensions have recently been proposed (e.g., Li and Racine 2008). In the presence of selection, there are, however, only a few approaches available to *point*, as opposed to set or partially, identify parameters of a quantile function – identification at infinity, the Buchinsky (1998) approach, and the Arellano and Bonhomme (forthcoming) approach. Olivetti and Petrongolo (2008) propose another approach but focusing only on median regressions. While they could slightly relax the assumption of selection on unobservables to impute wages for workers who work and have wages for more than a year, they still have to resort to the selection on observable assumption for those who never work. The first approach is based on the principle that selection bias tends to zero for individuals with certain characteristics who always work and whose probability to work is close to one (Heckman 1990; Mulligan and Rubinstein 2008; Chamberlain 1986). As a result, quantile functions can be identified using the selected sample (even in the absence of exclusion restrictions). However, the definition of "closeness" to one can be arbitrary in practice and there is a significant trade-off between sample size and the amount of selection bias. Mulligan and Rubinstein (2008) adopt this approach to assess the robustness of their conditional mean results. They define "closeness" to one as probability of working equal to or greater than .8, and the resulting sample is only about 300 observations per five-year sample, less than 1% of the original sample.

and Bonhomme (2016) note that it is unlikely to specify a data generating process consistent with the Buchinsky assumptions except in the case of either 1) additivity and parallel quantile curves, implying quantile functions are identical and equal to the conditional mean function, or 2) selection is random; see, also, Melly and Huber (2008). Arellano and Bonhomme (forthcoming) proposed copula apparoach is in the spirit of full information likelihood methods favored by some Cowels Foundation researchers, without IVs, but with greater flexibility in modelling the marginal variables. In the presence of selection, their approach entails shifting the percentiles as a function of the amount of selection.

Consider the following quantile wage function (see, e.g., Chernozhukov and Hansen 2008)

$$\ln(w) = g(x, u) \quad u|x \sim Uniform(0, 1) \tag{15}$$

where $\tau \mapsto g(x_i, \tau)$ is strictly increasing and continuous in τ. This can be a non-separable function of observable characteristics, x, and unobservable disturbances u, normalized and typically interpreted as ability (Doksum 1974; Chernozhukov and Hansen 2008).[22]

The labor force participation decision must be modelled and is determined by reservation wages:

$$\begin{aligned} S &= I(\ln(w) - R(z) - \eta \geq 0) \\ &= I(g(z) - \eta \geq 0) \end{aligned} \tag{16}$$

where $R(z) + \eta$ is the reservation wage (determined by both observable and unobservable characteristics). $I(\cdot)$ is an indicator function (equal to one if the argument is true, zero otherwise); thus, $S = 1$ if the wage offer is greater than or equal to reservation wage. Let $z = (x', \widetilde{z}')'$, where $\widetilde{z}$ includes a vector of IVs statistically independent of both (u, v) given x. An exclusion restriction is through a variable that affects reservation wages only, for example, the monetary value of leisure.

The last equation is commonly re-written as follows

$$S = I(v \leq p(z)) \quad v|x \sim Uniform(0, 1) \tag{17}$$

where $p(z) = \Pr[S = 1|z]$ is the propensity score, and assuming $p(z) > 0$ with probability one.[23]
In the presence of selection,

$$\Pr[\ln(w) \leq g(x, \tau)|s = 1, z] = \Pr[u \leq \tau | v \leq p(z), z] = \frac{C_x(\tau, p(z))}{p(z)} \equiv G_x(\tau, p(z)) \neq \tau$$

where the joint cumulative distribution function (or copula) of (u, v) is defined as $C_x(u, v)$. The observed rank for the τ^{th} quantile, $g(x, \tau)$, is no longer the τ in the selected sample. Instead, the observed rank is $G_x(\tau, p(z))$. Knowledge of the mapping between the quantile and its observed rank in the sample allows estimatation of $g(x, \tau)$ using a "rotated quantile regression". This is indeed the

[22] $\Pr[\ln(w) \leq g(x, \tau)|x] = \Pr[g(x, u) \leq g(x, \tau)|x] = \Pr[u \leq \tau|x] = \tau$. The first equality follows from Equation (15). The second follows from the fact that conditional on x, u is uniformly distributed.

[23] Note that Equation (17) is a normalization commonly used in the treatment effects literature. Note that $\mathbb{E}[S|z] = \Pr[S = 1|z] = p(z) = \mathbb{E}[S = 1|p(x)] = \Pr[S = 1|p(z)]$.

idea proposed by Arellano and Bonhomme (forthcoming). [24]

4.1.3. Practical Implementation

Propensity scores are estimated by probit models which include polynomial terms of the continuous variables up to third order, as well as the interaction terms between them and other discrete variables, in addition to the IV. These variables enter the probit model additively. A linear index model makes "identification and estimation simple and transparent." (Bonhomme et al. 2014)

We work with a linear conditional quantile function $g(x, u) = x'\beta(u)$. Despite the linearity for a given quantile it is still *nonlinear* in nature because this specification allows x to have differential impacts on the wage distributions.[25] And it is a non-separable function of x and u, allowing for interaction between the observable and unobservable characteristics, and is thus preferred to the additive structure that is often assumed in the conditional mean models. Moreover, Angrist et al. (2006) suggest that linear quantile regression provides a weighted least squares approximation to an unknown and potentially nonlinear conditional quantile regression. Below we provide some graphic evidence of the performance of such linear yet non-separable models.

While identification analysis in Arellano and Bonhomme (forthcoming) is general and covers the case where the copula is nonparametric, we adopt their choice of the Frank copula depending on a low-dimensional vector of parameters. This copula is widely used in empirical work (Meester and MacKay 1994; Trivedi and Zimmer 2005). its single parameter, ρ, captures dependence between $G_x(\tau, p(z)) \equiv G_x(\tau, p(z); \rho)$. Frank copula is "comprehensive" because it permits a wide range of potential dependencies, including negative dependence. It provides for a *finite level of data driven dependence.*

The dependence parameter ρ has an additional useful interpretation, indicating the sign of selection. A *negative* ρ indicates *positive* selection into employment, while *positive* ρ implies *negative* selection. This facilitates the comparison to the patterns of selection over time reported in the literature, e.g., Mulligan and Rubinstein (2008). ρ, is further allowed to be gender-specific (We also provide some checks that suggest this approach works quite well here with the CPS samples. See subsection "Results" below)

There are three-steps: Estimate propensity scores, $p(z)$; Estimate the dependence parameter, ρ; and given the estimated ρ and a specified τ, obtain the observed rank, $G_x(\tau, p(z); \rho)$ and estimate β_τ using the"rotated quantile regression". To recover the unconditional distribution, we estimate β_τ for $\tau = 0.02, 0.03, \ldots, 0.97, 0.98$.[26]

[24] The algorithm is provided in detail in Appendix B in the supplemental material. Exclusion restrictions and functional forms regarding $G(\cdot)$ provide identification.

[25] As noted in Melly and Huber (2011), "allowing for arbitrary heterogeneity and nonseparability" only identifies the bounds of the effects which are "usually very wide in typical applications".

[26] The third step is computationally intensive because, for each year of the data, a large number of quantile regressions must be estimated. Further, we conduct inferences based on 299 replications (which requires estimation of more than a million quantile regressions for the comparison of every two pairs of distributions.). The implementation details can be found in the supplemental material.

4.2. Variables and Exclusion Restrictions

The vector, x, is a typical set of wage determinants, including educational attainment dummies, marital status, polynomial terms of age up to third order, racial dummy and regional dummies. This is the common set of covariates in the literature on the gender gap with the CPS data. The corresponding wage equation is similar to what Blau and Kahn (2016) refer to as "human capital specification". [27]

There are two popular IVs for the participation equation, husband's income and the presence of young children (e.g., Mulligan and Rubinstein 2008; Machado 2012; Buchinsky 2001; Chang 2011; Martins 2001). For example, Mulligan and Rubinstein (2008) use the number of children younger than six, interacted with marital status as variables determining employment, but excluded from the wage equation.[28]

Here, we follow the tradition in this literature to use the presence of any young children under age 5 as our (overidentifying) excluded IV. To assess the validity of our exclusion restriction, we present two sets of results. First, the strength of empirical relationship between our IV and labor force participation decision (see estimates in Table (A.1) in the supplemental material). The number of young children indeed has a positive and statistically significant effect on labor force participation rates among women. The magnitude of the effect has been decreasing over time. Specifically, having one more young child can reduce female labor force participation rate by nearly 18 percent in 1976, but only 7 percent in 2013. By contrast, we fail to find robust evidence of young children effect on men's labor force participation. Men's working decisions may be motivated by different factors than women's. Note that, this exclusion restriction is not required for identification which is delivered by the copula. (We thank Jim Heckman for pointing this out.)

The second set of results is concerned with the independence of the excluded IV and potential wages (at least conditional on X). Huber and Mellace (2014) test the validity of such exclusion restrictions in this setting. They consider eight empirical applications and find husband's income is not a valid instrument, but the validity of the number of young children "is not refuted on statistical grounds". Using their method, we too fail to reject the validity of our IV in all years.[29] Results are presented in Table (A.2) in the supplemental material.

[27] Recent prominent examples in the field include Blau and Kahn (2006) and Mulligan and Rubinstein (2008). Buchinsky (1998) uses a similar set of variables to estimate the conditional quantile regressions for women in the presence of selection. Card and DiNardo (2002) and Juhn and Murphy (1997) also employ a similar set of wage determinants to study wage inequality. As noted in Buchinsky (1998), other data sets such as the PSID (Panel Study of Income Dynamics) and the NLS (National Longitudinal Survey) may contain a potentially richer set of variables, "but suffer from other problems (such as attrition)." It is thus difficult to be certain whether different findings, if any, are due to differences in the control set, or the differences in data design, or representativeness of the population. Note that Olivetti and Petrongolo (2008), although using the PSID, employs a similar set of covariates as in our work.

[28] Also noted in Machado (2012), the number of children is used as an explanatory variable in the shadow price function in Heckman (1974), "one of the seminal works on female selection", and an IV in the participation equation in Heckman (1980). The number of young children may affect women's reservation wages and their labor supply decisions because it could affect "the value of leisure" for women (Keane et al. 2011) and child-rearing is time consuming and costly. On the other hand, whether husband's income can theoretically affect women's labor force participation is debatable. For example, Keane et al. (2011) notes that the linearity and separability of consumption in the utility function implies that husband's income does not affect women's labor force participation decision

[29] They show that under our model assumptions, the following inequalities hold

$$\begin{aligned}\mathbb{E}[\ln(w)|\widetilde{z}=1, S=1, \ln(w)\leq y_q] &\leq \mathbb{E}[Y|\widetilde{z}=0, S=1]\\ &\leq \mathbb{E}[\ln(w)|\widetilde{z}=1, S=1, \ln(w)\geq y_{1-q}]\end{aligned}$$

4.3. Results

Firstly, we may assess how well the quantile selection models perform under these assumptions. We are able to identify and compare the wage distribution of the full-time workers based on quantile selection model, with the observed wage distribution in our sample. Figure (A.1) in the supplemental material displays this comparison at a few quantiles, $\tau = .10, .25, .50, .75, .90$.

The quantile models perform reasonably well. In most years, specification errors are within a very small neighborhood. In some instances the imputed quantiles are identical or close to identical to the observed ones. While this is not a formal test of all the assumptions, it provides some confidence in the methods used here and the results that follow. The "derived" mean and median gap measures are also close to the corresponding ones obtained by Mulligan and Rubinstein (2008) and Olivetti and Petrongolo (2008) using completely different approaches when addressing selection for women.[30]

4.3.1. Selection and The Magnitude of the Gap

The estimated dependence parameter, ρ, for women and men are presented in Tables (4) and (5), respectively. It varies in magnitude and direction over time, and by gender. For women, it is mostly positive up to 1991, close to zero in 1992 and increasingly negative from then on. Positive dependence indicates negative selection, while negative dependence suggests positive selection. This pattern is consistent with Heckman (1980)'s early finding (the non-working women are often the high-wage women) and Mulligan and Rubinstein (2008).[31] Neal (2004) similarly emphasizes that the selection pattern can be either positive or negative. Below we will discuss two reasons for the observed transition in the selection pattern for women and formally test one of them. For men, on the other hand, the parameter is consistently negative throughout the same period, except for very few early years, which suggests positive selection.

Addressing selection for women would lead to a smaller gender gap in the presence of negative selection, but a larger gender gap in the presence of positive selection. The observed transition in the selection pattern for women implies a smaller convergence than suggested by the literature. Indeed, addressing selection only for women, Mulligan and Rubinstein (2008) conclude that the gender gap may have not shrunk at all, indicating that women continued to fare worse than men during their study period. In an earlier version of our paper, Maasoumi and Wang (2014), we indeed find strikingly similar results not only in terms of patterns but also of magnitudes.[32] However, the direction of the

---

Such inequalities imply the following null hypotheses:

$$\begin{aligned} \mathbb{E}[\ln(w)|\widetilde{z}=1, S=1, \ln(w) \le y_q] - \mathbb{E}[\ln(w)|\widetilde{z}=0, S=1] &\le 0 \\ \mathbb{E}[\ln(w)|\widetilde{z}=0, S=1] - \mathbb{E}[\ln(w)|\widetilde{z}=1, S=1, \ln(w) \ge y_{1-q}] &\le 0 \end{aligned}$$

Huber and Mellace (2014) propose a test procedure to verify these inequalities. Note that this test can be readily extended to the multivalued case, but for ease of exposition and computation, we consider a binary case here, i.e., the presence of young children. A negative test statistic with a large p-value indicates that IV validity is *not* violated. Readers who are interested in the details of this procedure are referred to Huber and Mellace (2014)

[30] We note that the errors are larger for three percentiles ($\tau = 0.10, 0.75, .90$) for the year 2010, which is probably due to the fact that some models for the year 2010 have difficulty converging. We advise caution with results for 2010. Unless otherwise noted, we have excluded 2010 from our analysis.

[31] We thank Jim Heckman for pointing this out.

[32] Although using completely different approaches, our early estimates of both the mean and median gaps (addressing the selection only for women) are strikingly similar to what is found in Mulligan and Rubinstein (2008) and Olivetti and

Table 4: Dynamics of Selection Parameter and Its Signs (Female)

Year	ρ	Sign	Year	ρ	Sign	Year	ρ	Sign	Year	ρ	Sign
1976	2.164	N	1986	0.602	N	1996	-1.098	P	2006	-1.098	P
1977	1.747	N	1987	-0.241	P	1997	-1.098	P	2007	-1.548	P
1978	1.034	N	1988	0.360	N	1998	-0.664	P	2008	-1.884	P
1979	0.909	N	1989	0.120	N	1999	-0.543	P	2009	-1.953	P
1980	1.416	N	1990	-0.603	P	2000	-0.726	P	2010	0.060	N
1981	1.416	N	1991	0.240	N	2001	-0.181	P	2011	-2.534	P
1982	1.097	N	1992	-0.001	P	2002	-1.098	P	2012	-3.539	P
1983	0.542	N	1993	-0.061	P	2003	-1.748	P	2013	-2.688	P
1984	1.097	N	1994	-0.421	P	2004	-1.614	P			
1985	0.663	N	1995	-0.603	P	2005	-1.482	P			

Table 5: Dynamics of Selection Parameter and Its Signs (Male)

Year	ρ	Sign	Year	ρ	Sign	Year	ρ	Sign	Year	ρ	Sign
1976	0.725	N	1986	-0.482	P	1996	-0.726	P	2006	0.120	N
1977	-0.121	P	1987	-0.181	P	1997	-2.534	P	2007	-0.543	P
1978	0.300	N	1988	-1.224	P	1998	-6.339	P	2008	-2.688	P
1979	1.034	N	1989	-2.767	P	1999	-2.383	P	2009	-4.029	P
1980	0.848	N	1990	-2.534	P	2000	-1.548	P	2010	-6.528	P
1981	0.972	N	1991	-2.165	P	2001	-1.482	P	2011	-5.087	P
1982	-0.301	P	1992	-1.288	P	2002	-1.161	P	2012	-4.827	P
1983	-2.023	P	1993	-3.927	P	2003	-1.816	P	2013	-2.093	P
1984	-3.356	P	1994	-5.663	P	2004	-1.748	P			
1985	-0.001	P	1995	-5.223	P	2005	-1.161	P			

changes is unclear, a priori, when selection is accounted for *both* men *and* women.

Results are presented in Table (6). Corresponding standard errors are provided in Table (C.30) in the supplemental material. Regardless of measure, the adjusted gender gap (accounting for selection) is different from the baseline. The differences are substantial, especially for those in the lower half (including the medians) of the distribution, and in summary measures such as the mean and entropy measures. In early years, for example, 1976 and 1977, the gap is smaller than the gap in fully employed sample, regardless of the measure. The difference can be as large as 16 percentage points at the 10^{th} percentile, implying that the gap is overestimated by roughly 38 percent by baseline estimates. On the other hand, in many later years, the gap is much greater when selection bias is removed.

We find that both the signs and magnitudes of the gap change at certain percentiles. While a positive gap persists at the upper tail of the distribution between men and women, the gap sometimes becomes even negative in the lower tail, indicating that low-wage women do not necessarily always perform worse than low-wage men. This result is masked by simple examination of mean or median. It also implies that there would be no stochastic dominance ranking, making the choice of a summary measure both necessary and sensitive. This is indeed the reason why we introduce and prefer a measure like S_ρ, which accounts for all the moments of the distribution as well as inequality/dispersion.

We find that the gender gap is larger in the upper tail than in the lower tail of the wage distributions. Regardless of the measure, the gender gap accounting for selection also evolves very differently compared to the baseline. The gender gap is much more cyclical than the baseline estimates and fluctuates even more in the lower tail of the distribution. These results are summarized in Figure (3) to be compared with the baseline case in Figure (1).

We plot the smoothed trend line for each measure in Figure (2) above. Fluctuations notwithstanding, there appears to be three distinct phases, especially in the lower tails, in contrast to the two-phase pattern noted without accounting for selection. We can now see a rapidly declining trend in early years ("fast convergence" period), then a period of stagnant growth or even a reversal in the trend, followed by a further declining trend since the great recession.

The first phase is similar to that observed for the full-time employed sample, but with different magnitudes. The differences vary across measures, and can be seen by comparing Table (7) to Table (2). The "true" gender gap converges at slower rates in the upper tails, but convergence is more pronounced in the lower tail. The second phase of the trend is even more distinct from that implied by the selected sample. Not only is the gap more stagnant in the upper tail of the distribution; there is a reversal of the trend in the lower tail. This trend appears to continue until the most recent recession.

The pattern since 2007, the start of the recent recession, is worth noting. While there is a lack

---

Petrongolo (2008) (that focus on only mean and median). Using the CPS data from 1975-2001, Mulligan and Rubinstein (2008) find that the raw gender gap without addressing the selection issue is 0.419 in 1975-1979 and .256 in 1995-1999. These estimates are close to ours presented in the previous working paper. After correcting for the selection using the Heckman selection model, they find that the mean gender gap was -0.379 in 1975-1979 and -0.358 in 1995-1999, similar to our results ranging from -.321 to -.393 in 1975-1979 and from -.333 to -.374 in 1995-1999. Our median results are also similar to what is found in Olivetti and Petrongolo (2008) using the PSID data from 1994-2001. For example, their results using the imputation method based on wage observations from adjacent waves range from .339 to .363 (in their Table 2), and the results using the imputation method based on observables from a probabilistic model range from .359 and .371. These estimates are similar to our results during the same period, ranging from .330 to .384. The fact that these results are also in line with the literature provides some confidence in our results for the rest of the distribution.

Table 6: Measures of The Gender Gap (with selection correction)

Year	S_ρ (1)	Theil (2)	Mean (3)	10th (4)	25th (5)	50th (6)	75th (7)	90th (8)
1976	6.712	14.449	0.294	0.200	0.274	0.318	0.335	0.344
1977	6.919	16.502	0.282	0.136	0.239	0.315	0.354	0.376
1978	9.400	23.489	0.357	0.210	0.315	0.391	0.425	0.432
1979	10.432	23.824	0.389	0.249	0.343	0.427	0.459	0.460
1980	8.712	19.876	0.344	0.198	0.299	0.384	0.420	0.422
1981	8.074	18.028	0.330	0.186	0.280	0.364	0.404	0.412
1982	7.288	18.107	0.302	0.140	0.239	0.340	0.383	0.397
1983	5.396	13.219	0.231	0.025	0.141	0.268	0.342	0.375
1984	3.706	8.842	0.107	-0.118	-0.001	0.134	0.235	0.289
1985	6.075	16.370	0.294	0.142	0.231	0.325	0.375	0.397
1986	5.198	11.870	0.273	0.114	0.213	0.308	0.360	0.367
1987	5.700	11.784	0.314	0.185	0.269	0.354	0.384	0.379
1988	3.986	9.028	0.233	0.086	0.166	0.255	0.319	0.338
1989	3.365	8.576	0.189	0.033	0.124	0.216	0.274	0.304
1990	3.603	7.692	0.239	0.098	0.187	0.266	0.315	0.328
1991	2.487	5.446	0.171	0.030	0.110	0.192	0.246	0.277
1992	2.937	6.931	0.216	0.088	0.175	0.242	0.281	0.295
1993	2.233	5.196	0.071	-0.120	-0.015	0.095	0.179	0.226
1994	2.554	5.835	0.039	-0.196	-0.078	0.071	0.176	0.228
1995	2.163	4.583	0.079	-0.131	-0.011	0.113	0.196	0.235
1996	3.227	9.438	0.262	0.181	0.233	0.280	0.298	0.302
1997	2.562	3.869	0.218	0.109	0.181	0.237	0.270	0.282
1998	2.863	7.682	0.044	-0.206	-0.060	0.079	0.173	0.222
1999	2.134	5.460	0.176	0.040	0.124	0.189	0.236	0.268
2000	2.370	4.673	0.216	0.116	0.178	0.232	0.262	0.282
2001	2.010	3.297	0.187	0.092	0.140	0.190	0.225	0.270
2002	2.798	8.009	0.246	0.167	0.209	0.248	0.276	0.312
2003	2.673	6.918	0.241	0.142	0.212	0.248	0.277	0.312
2004	2.190	4.013	0.227	0.139	0.196	0.237	0.262	0.297
2005	2.554	5.665	0.246	0.175	0.217	0.249	0.274	0.302
2006	3.281	9.995	0.284	0.230	0.258	0.284	0.299	0.332
2007	3.148	6.082	0.283	0.241	0.260	0.283	0.298	0.325
2008	1.886	3.137	0.208	0.141	0.181	0.209	0.236	0.260
2009	1.743	3.329	0.145	-0.013	0.082	0.168	0.216	0.259
2010	18.815	-7.687	-0.115	-0.627	-0.454	-0.079	0.140	0.339
2011	1.249	2.625	0.094	-0.059	0.013	0.108	0.181	0.231
2012	1.581	3.749	0.161	0.010	0.096	0.181	0.237	0.277
2013	2.352	5.081	0.259	0.186	0.236	0.273	0.292	0.309

[1] Note that S_ρ and Theil are multiplied by 100.

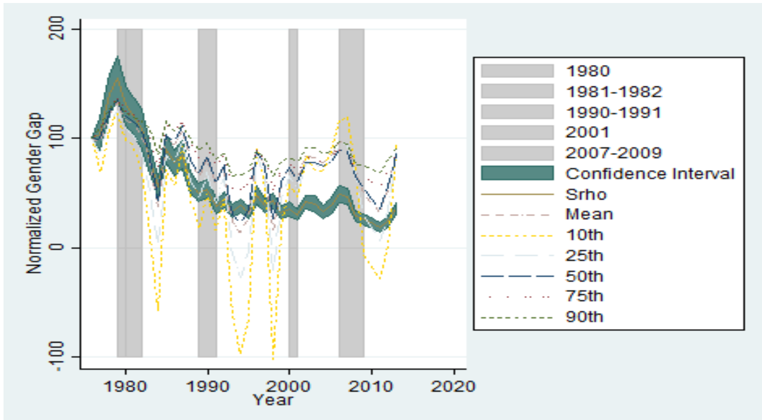


Figure 3: The Trend of Selection-corrected Gender Wage Gap (Shaded Areas correspond to the recession periods announced by NBER)

Table 7: Implied Long-Run Annual Changes in the Gender Gap with selection corrected

Period	S_ρ (1)	Theil (2)	Mean (3)	10th (4)	25th (5)	50th (6)	75th (7)	90th (8)
1973-2013	-0.042	-0.045	-0.018	-0.039	-0.025	-0.018	-0.014	-0.011
1973-1994	-0.053	-0.053	-0.032	-0.091	-0.048	-0.031	-0.022	-0.018
1994-2006	-0.025	-0.027	0.012	0.104	0.029	0.006	-0.002	0.000
2007-2013	-0.038	-0.049	-0.026	-0.109	-0.047	-0.020	-0.010	-0.008

[1] These values are long-run compound annual change rates implied by the initial and the last smoothed values of each period. See footnote (**??**) as well.

of convergence in the upper tail, a general decreasing trend in the lower tail is observable. (This result is also different from those in Maasoumi and Wang (2014) where we address selection only for women). The gap soared from the beginning of the recession, while the full-time working sample gap remained relatively stable. Recall that both men and women who stay in full-time employment tend to be higher-wage earners during this period (positive selection). Such a difference indicates that while the recent recession may have hurt the labor market prospects of both low-skilled women and men, and "forced" them out of full-time employment, low-skilled men were probably even more severely impacted during the recent recession. This result is consistent with the fact that industries such as construction and manufacturing where low-skilled men are primary workforce were more impacted during the recession Sahin et al., 2010). In general, the selected (working) sample underestimates the gender gap, while addressing selection only for women exaggerates it.

The quantile specific observations are informative but fall short in representing the overall gender gap. We find below that there is no stochastic dominance ranking in quite a few periods, making the choice of a summary measure both necessary and sensitive. Both entropy measures suggest that, although there was some narrowing trend of the gap in early 1990s, such progress was much smaller than previously found in the literature that did not account for women's selection into employment. Also, the commonly found "slower" convergence (in the wage distributions between men and women) in later years is even slower.

A three-phase trend is evidenced: there was initially a strong convergence in the gender gap, the extent of which can be either under- or over-estimated when using the sample of full-time workers (depending on measures). This time trend, however, became plateaued, or even reversed, in the middle of this period (depending on the measures). During the most recent recession, there is some decline in the gap among low-wage workers, which is likely due to a worsened situation among low-skilled men.

4.3.2. Stochastic Dominance

A positive selection for women may strengthen the earlier findings of SD relations between the full time employed. Negative selection, on the other hand, may imply a crossing of distributions at lower wages.[33] Dominance relations are then less likely once selection for women is controlled for. The impact of addressing selection for *both* men and women requires further attention.

Table (8) summarizes the outcome of statistical SD tests, corrected for selection. Given the relatively large decline in the gender gap during the early years, it is less likely to find statistically significant first-order dominance, or any higher order dominance relations before the early 1990s. There is no statistical dominance relations in 4 cases (1977, 1986, 1990, and 1992), for which FSD is at low degree of confidence, less than .90), and none in 6 cases (1983, 1984, 1989, 1991, 1993, and 1994). The inability to rank order the earnings distributions between men and women in this case is equally informative. It implies that for all these years, say, 1994, a suggestion that women are worse off than men is not robust. Our entropy measure of the gap may be preferred in such situations when distributions cross, especially when they cross at lower wages.

[33] Negative selection implies that extremely high wage earners drop out full-time employment. Once they are included in the sample, it is more likely that women's wages at lower tail may be higher than before, and as a result, we may be more likely to observe crossing at the lower tail.

Table 8: Stochastic Dominance Results with Correction for Selection(Female v.s. Male Wage Distributions)

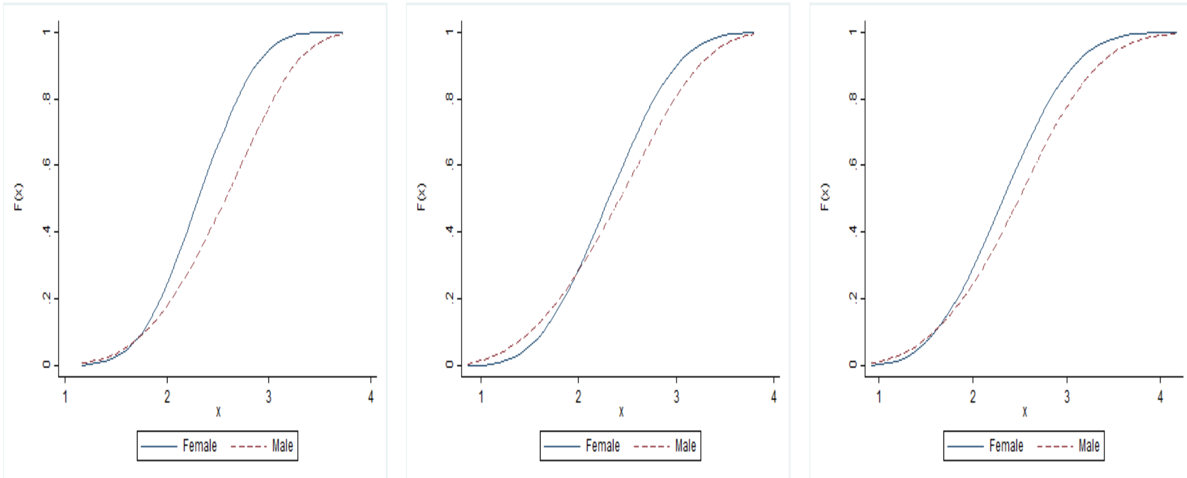
Year	SD	d	$Pr[d \leq 0]$	s	$Pr[s \leq 0]$	Year	SD	d	$Pr[d \leq 0]$	s	$Pr[s \leq 0]$
1976	F	-6.99	1.00	-6.99	1.00	1995	N	59.51	0.00	540.75	0.00
1977	F	-7.74	0.84	-7.74	0.84	1996	F	-6.81	1.00	-6.81	1.00
1978	F	-7.34	1.00	-7.34	1.00	1997	F	-7.16	0.51	-14.06	0.51
1979	F	-8.02	1.00	-11.21	1.00	1998	N	84.45	0.00	606.26	0.00
1980	F	-11.22	1.00	-14.72	1.00	1999	N	7.09	0.62	17.18	0.62
1981	F	-10.33	1.00	-10.33	1.00	2000	F	-7.83	1.00	-13.44	1.00
1982	F	-7.70	0.99	-7.71	0.99	2001	F	-8.85	0.99	-14.69	0.99
1983	N	16.00	0.00	89.98	0.00	2002	F	-9.13	1.00	-12.60	1.00
1984	N	57.25	0.00	514.90	0.00	2003	F	-9.50	1.00	-14.81	1.00
1985	F	-7.45	1.00	-11.23	1.00	2004	F	-8.99	1.00	-8.99	1.00
1986	F	-8.98	0.87	-13.20	0.87	2005	F	-9.56	1.00	-11.61	1.00
1987	F	-7.65	1.00	-10.73	1.00	2006	F	-9.52	1.00	-11.95	1.00
1988	F	-7.68	0.94	-9.78	0.94	2007	F	-9.69	1.00	-14.44	1.00
1989	N	16.18	0.00	86.12	0.00	2008	F	-10.38	1.00	-14.09	1.00
1990	F	-9.70	0.73	-9.70	0.73	2009	N	25.65	0.00	146.92	0.00
1991	N	11.17	0.04	48.41	0.04	2010	S	293.55	0.00	-14.36	1.00
1992	F	-8.57	0.70	-15.57	0.70	2011	N	36.73	0.00	284.95	0.00
1993	N	58.97	0.00	537.95	0.00	2012	N	16.55	0.23	63.72	0.23
1994	N	85.15	0.00	678.50	0.00	2013	F	-10.91	1.00	-11.95	1.00

[1] Only the d and s statistics and corresponding p-values are reported. **N** denotes no (first- or second-order) dominance found, while **F** denotes First-order dominance and **S** Second-order dominance.

In the period beyond the early 90s (except for 2010) men's earnings first-order dominate women's in the majority of the cases to a high degree of statistical confidence, while we are less likely to find dominance during the most recent recession.

Absent these tests, the mere observation of the gap at some percentiles may be inconclusive. For example, we find positive differences in wages in favor of men at all select percentiles in 1983 but fail to find any dominance relations. We plot the CDF comparisons for select years in Figure (4) and the results for other years are available in the supplemental material.

These graphs are illuminating. For all the non-dominance cases, there is an early crossing of the CDFs, while the CDF of men's wage distribution lies mostly under that of women's elsewhere. At the extreme lower tail, women perform better than men, while other women fare worse than men. This result is indeed the motivation for why we adopt our entropy measure and the SD approach to study the gender gap. Together, it illustrates and underscores the benefit to considering the *entire* distribution within a decision theoretic framework, and it also highlights what could be missed should we simply look at select parts of the wage distributions. A narrower class of preference functions would order these distributions. These must be "increasingly averse" to inequality at lower or higher ends of the earnings distribution. Indeed one can see how an "upward" aversion to inequality, as described recently in Aaberge et al. (2013) may rank these crossing distributions. The class of functions that may uniformly rank distributions that cross entail narrower and increasingly "non-consensus"



(a) 1983 (b) 1995 (c) 2009

Figure 4: CDF Comparisons of Female (Selection Corrected) and Male Wage Distributions For Select Years

interpersonal comparisons of well being, such as embedded in the Gini type of preference functions. In such situations it is more than usually important to be explicit about the properties of any evaluation function employed to characterize the gap and make decision on it.

5. Further Explorations by Educational or Racial Groups

We explore the gender gap by education and race.[34] Given constraints of space, we first summarize the common findings in these further explorations and contrast them to our full-sample results here. We then highlight some of the important differences across groups in each subsection below. Note that for such sub-group analysis, it is even more important to have summary measures such as our preferred entropy measures to summarize the within-group heterogeneity in the gender gap. Some important results are highlighted and summarized in the paper, but many actual estimates are presented in the supplemental material in the interest of space.

Consistent with our results above, regardless of educational level or race, the gender gap within each group is larger in the upper tail than in the lower tail of the wage distributions. We also find that the long-run trends of the gap for each group are surprisingly similar to the results using the full sample. Specifically, we again find slower convergence in the gap (than what is implied without addressing selection) or even lack of convergence in some cases. A distinct three-phase trend of the gap again exists among some educational or racial groups.

5.1. By Education

We repeat our analysis for four different educational groups (below high school education, high school, some college, and college and above). The actual estimates without selection correction are reported in Tables (C.6)-(C.9) in the supplemental material, and those with selection correction (C.10)-(C.13) . The gender gap among the least educated workers were larger than the gender gap in the rest

[34]We thank Jim Heckman for the suggestions that eventually led to these further results.

of the population. Blau (1998) examines the trends in well-being of American women by educational groups during the period 1970-1995. One important findings of her paper is "the deteriorating relative economic position of less educated women." Evaluation of such statement depends on the choice of benchmark (which, in this case, is men with similar educational levels) and the choice of summary measures. In the interest of space, we highlight the differences in the implied long-run changes for each educational group characterized by both entropy measures in Table (9), and the corresponding graphs are in Figure (C.3) in the supplemental material.

We first look at the results without addressing selection in Panel A of Table (9). The findings are generally similar across educational groups. The implied convergence in the gap was notable before the mid-1990s, and continued at a somewhat modest rate. The starting levels of the gender gap and the implied rates of decline are different across groups and time periods. Until early 1990s, the extent of the convergence in the overall gap (measured by S_ρ) is larger among the less educated workers (those with high-school and below high school education), whereas the declines are relatively smaller for workers with some college education, and smallest for workers with more than 4-year college education. Over this period, the average annual percentage changes are -6.2 and -5.9 percents for workers with high-school and below, respectively. The implied annual percentage changes are -5.4 percent for workers with some college education and only -4.3 percent for those with more than 4-year college education. Afterward, the progress among the least educated workers stagnated, while women in other groups continued to narrow the gender gap with their male counterparts. Specifically, the implied annual percentage changes were only -1.7 percent, while the average annual percentage changes were about -2.7 percent for college graduates. This pattern is very similar to what is found in Blau (1998).

Table 9: IMPLIED LONG-RUN ANNUAL CHANGES IN THE GENDER GAP BY EDUCATION

	S_ρ				Theil			
Period	Less than High School (1)	High School (2)	Some College (3)	College & Above (4)	Less than High School (5)	High School (6)	Some College (7)	College & Above (8)
Panel A: Without Selection Correction								
1973-2013	-0.040	-0.046	-0.042	-0.036	-0.044	-0.047	-0.043	-0.039
1973-1994	-0.059	-0.062	-0.054	-0.043	-0.061	-0.064	-0.055	-0.046
1994-2013	-0.017	-0.026	-0.028	-0.027	-0.023	-0.027	-0.028	-0.029
Panel B: With Selection Correction								
1973-2013	-0.019	-0.027	-0.026	-0.027	-0.018	-0.037	-0.032	-0.045
1973-1994	-0.038	-0.044	-0.038	-0.041	-0.034	-0.056	-0.037	-0.035
1994-2006	0.004	-0.008	-0.005	-0.002	0.001	-0.018	-0.027	-0.026
2007-2013	-0.004	-0.011	-0.023	-0.025	-0.007	-0.017	-0.029	-0.102

[1] These values are long-run compound annual change rates implied by the initial and the last smoothed values of each period.

As noted in Blau (1998), one interesting question is whether the declining relative wages of the least educated is simply a result of compositional changes within the group. To shed light on this issue, we turn to our results when the selection issue is addressed (Panel B of Table 9). Initially, the gap among the least educated workers exhibit a slower convergence over time, compared to other educational groups. During the second phase, the gap among the least educated workers even increased slightly, while there was still slight convergence among other educational groups. During the most recent recession, while women in other educational groups continue to catch up with their male counterparts at relatively fast rates, the progress is much smaller among women with least education. Specifically, the implied rates of convergence are only 0.4 percent for women with least education, but 2.5 percent among women with college education and above, the largest among all educational groups. Our results indicate that "the deteriorating relative economic position of less educated women" over this period could be well underestimated.

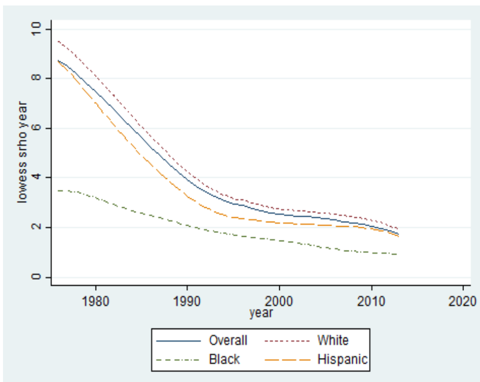
Turning to SD tests (those without selection correction are reported in Tables (A.3)-(A.6) in the supplemental material, and those with selection correction (A.7)-(A.10)), we find that when failing to address selection, the results for individuals with more than college education, some college, or high school education are very similar to the full-sample results above, and we generally find statistically significant, first-order dominance relations. Further, although generally statistically insignificant, we also observe dominance relations among the least educated individuals. This result again confirms that, without controlling for the selection issue, women in all educational groups fare worse compared to their male counterparts, despite the convergence in the past decades. The results change, however, when taking into account non-full-time workers. For individuals without college degrees (less than high school education, high school education, or some college education), we find that while there existed dominance relations in the first few years, the occurrence of such relations decreased drastically during the period of convergence, but increased again in the later years (beyond 1994). On the other hand, despite some convergence, we find dominance relations in nearly all years for individuals with more than college education. This result implies that women with college education fare worse than their male counterparts, while women with less education do not necessarily. This later result again demonstrates a certain inevitability of subjective Evaluative Functions when distributions cross at low wages and cannot be ordered by first or second order stochastic dominance.

5.2. By Race

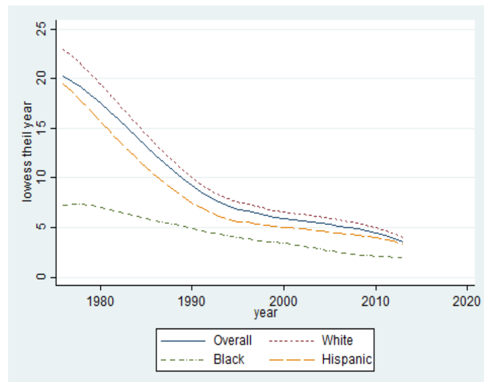
The gap by race is reported in Tables (A.11)-(A.13) in the supplemental material, and those with selection correction (A.14)-(A.16). Addressing selection has different impacts on the gap measures across race, time, and the wage distribution. Most measures for whites understate the gap in the majority of later years (beyond early 1990s), while mostly overestimate in early years. This pattern is similarly observed for the entropy measures. For Hispanics, the actual gender gap is generally underestimated, by mean and median. For blacks, the patterns are quantile dependent. Specifically, the actual gender gap is generally underestimated in the upper tail of the distribution, but less likely so in the lower tail. The latter result suggests that the gap among lower-skilled workers is generally smaller than for full-time workers. Failure to take into account these individuals may therefore overestimate the gap in the lower tail of the wage distribution. Indeed, once selection is accounted for, there are quite a few years in which low-skilled women actually performed better than their male counterparts. Taking into account the differences across the entire distribution and inequality within the group, the entropy measures imply that the gender gap is generally underestimated for both Hispanics and blacks as a whole.

The gender gap is larger in the upper tail than in the lower tail. The magnitude of the gap across racial groups is ordered. Specifically, the gap among minority groups (blacks and Hispanics) is smaller than amongst whites. Addressing selection, we find that the gap amongst blacks is also generally smaller than the gap amongst Hispanics.

Despite similar trends, black women as a whole have made much smaller progress in narrowing the gap, compared to both Hispanics and whites (as also evident from the implied trends of entropy measures in Figure (5)).



(a) S_ρ



(b) Theil

Figure 5: Comparison of Smoothed Trend of The Gender Gap with Selection Correction By race (Excluding 2010)

Turning to our SD results without addressing selection (those without selection correction by race are reported in Tables (C.14)-(C.16) in the supplemental material, and those with selection correction (C.17)-(C.19)), we generally observe first-order dominance relations, regardless of race. While such dominance relations are statistically significant for most years among whites and Hispanics, they are not (for many years) amongst blacks. Once selection is accounted for, we fail to observe dominance relations in considerably more years. For example, based on full-time workers, the wage distribution of white males first-order dominates that of white females in all years, but in only about two thirds of the years when selection is accounted for. These results are again due to crossing in the distributions at low wages; low skilled women do not necessarily perform worse than men in the labor market, consistent with our observations from conventional gap measures.

6. A New Concept of the Gender Gap: Taking into Account Value of Time

Our discussion has thus far focused on comparisons of the distributions of men's and women's *potential wage offers*. Our purpose in comparing these distributions is to evaluate relative well-being. Comparison of wage offers, however, does not fully serve this purpose. As we have seen, the presence of young children reduces the probability of a woman being a full-time worker. And as the literature has also noted, "a noteworthy number of these women are married to men who earn relatively high incomes" (Neal, 2004). For individuals, especially women, who do not work full-time, their decisions to stay home do not necessarily reflect low wage offers, but rather "high shadow prices of time spent at home" (Neal, 2004). In other words, wage offers do not necessarily represent income levels that they may enjoy, or the well-being of those who do not work full-time or work at all. To take into account the non-market value of time, an interesting yet useful comparison would then be based on an alternative wage distribution for men and women, defined by the mixed distribution of wage offers for workers (conditional on full-time employment) *and* reservation wages for non-working individuals (conditional on non-full-time employment), instead of the distribution of wage offers for all individuals.

Our quantile approach allows such analysis. In our framework, the actual monetary value of not working is captured by reservation wages. We can recover the distribution of reservation wages given unemployment by exploiting the selection equations (16) and (17). This involves a three-step procedure. First, having estimated our selection equation (17), we can obtain $\widehat{g(z)}$, the difference between the potential wage offers and reservation wages. Second, we can then draw potential wages, $\ln(w) = x'\beta(u)$ given non-full-time employment ($S = 0$) for given x and estimated coefficients, $\widehat{\beta}$. Finally, it is straightforward to recover the distribution of reservation wages in the third step. In fact the quantiles of this distribution depends on β and the copula of (U, V). In a different context, Bonhomme et al. (2014) rely on the selection equation to recover the distribution of agents' underlying preferences in a similar way.[35]

Various gender gap measures and SD tests are presented in Tables (10) and (11). We first find that the gender gap is much smaller between men and women when taking into account the reservation wages. This is not surprising because, conditional on non-full-time employment, the distribution of reservation wages should be in general better than the distribution of potential wage offers. And indeed, we find that in earlier years, considering the actual monetary benefits (captured by their reservation wages for non-full-time workers), women in the upper tail of distribution actually perform even better than men. This is consistent with our finding of negative selection in early years that women who do not work full time are generally high wage earners. This finding is emphasized by SD tests in Table (11). We observe even fewer instances of either first- or second-order dominance relations.

However, while women continue to perform better than men in the lower tail of the mixed distribution over time, the relative performance of women to men has changed in the rest of the distribution, with the gap widening in the upper tail. This is also evident with the smoothed trend of the gap at each part of the distribution in Figure (6). The smoothed trends of the gender gap at both 75^{th} and 90^{th} percentiles exhibit an increasing trend, while those of the gap at 10^{th}, 25^{th} and 50^{th} percentiles showing a S-shape trend with the gap first decreasing, then increasing before decreasing again in the most recession. Confirming this result, the entropy gender gap exhibits similar patterns to the gaps in the lower half of the distribution, while the mean gap is more consistent with the upper half of the distribution. This result implies that when taking into account value of time, women's relative well-being, especially among those in the upper tail, may have worsened over time; the extent of the deteriorating situation could be more severe than the traditional approaches suggest. This finding also highlights the importance of taking into account the value of time in the analysis of the gender gap.

[35] This entire subsection was indeed inspired by helpful discussions with Stephane Bonhomme, to whom we are grateful.

Table 10: Measures of The Gender Gap between the Mixed Distributions of Market and Non-Market Values

Year	S_ρ (1)	Theil (2)	Mean (3)	10th (4)	25th (5)	50th (6)	75th (7)	90th (8)
1976	1.860	4.722	-0.130	-0.118	-0.080	-0.068	-0.132	-0.242
1977	1.455	3.452	-0.111	-0.144	-0.088	-0.050	-0.078	-0.175
1978	0.953	5.085	-0.018	-0.061	0.004	0.045	0.018	-0.080
1979	1.157	2.769	0.033	-0.016	0.054	0.103	0.077	-0.028
1980	1.117	3.009	0.020	-0.045	0.033	0.088	0.073	-0.024
1981	1.121	7.012	0.009	-0.049	0.019	0.073	0.058	-0.033
1982	0.851	1.928	-0.033	-0.087	-0.026	0.020	0.017	-0.065
1983	0.929	1.694	-0.099	-0.156	-0.114	-0.071	-0.047	-0.086
1984	2.495	5.336	-0.200	-0.289	-0.239	-0.171	-0.121	-0.155
1985	0.577	1.208	0.004	-0.064	-0.011	0.045	0.064	0.013
1986	0.787	1.685	-0.011	-0.092	-0.025	0.038	0.056	-0.005
1987	0.684	1.390	0.046	-0.024	0.033	0.089	0.108	0.053
1988	0.665	1.409	0.008	-0.104	-0.040	0.037	0.090	0.078
1989	0.849	1.750	-0.024	-0.153	-0.085	-0.001	0.063	0.077
1990	0.833	1.814	0.031	-0.098	-0.020	0.062	0.115	0.117
1991	0.902	2.084	-0.033	-0.159	-0.087	-0.006	0.053	0.062
1992	0.661	1.536	0.019	-0.089	-0.020	0.047	0.091	0.092
1993	1.999	4.470	-0.124	-0.292	-0.217	-0.111	-0.014	0.029
1994	2.834	7.518	-0.153	-0.366	-0.274	-0.138	-0.018	0.042
1995	2.114	4.982	-0.103	-0.297	-0.215	-0.086	0.022	0.074
1996	0.907	1.908	0.090	-0.019	0.040	0.104	0.154	0.170
1997	1.005	2.383	0.062	-0.060	-0.002	0.076	0.138	0.167
1998	3.030	9.046	-0.104	-0.354	-0.237	-0.087	0.042	0.112
1999	1.318	2.945	0.033	-0.133	-0.048	0.044	0.123	0.166
2000	1.179	2.570	0.090	-0.045	0.032	0.107	0.163	0.194
2001	1.083	2.613	0.060	-0.069	-0.007	0.065	0.123	0.175
2002	1.366	3.283	0.116	-0.009	0.057	0.122	0.175	0.223
2003	1.161	2.568	0.103	-0.019	0.047	0.110	0.163	0.212
2004	1.007	2.520	0.083	-0.032	0.029	0.091	0.142	0.189
2005	1.084	2.580	0.096	-0.010	0.036	0.097	0.151	0.199
2006	1.562	4.131	0.143	0.031	0.091	0.150	0.196	0.245
2007	1.491	3.534	0.151	0.045	0.101	0.154	0.204	0.247
2008	1.018	2.568	0.078	-0.030	0.017	0.075	0.137	0.189
2009	1.487	3.687	0.012	-0.164	-0.085	0.016	0.104	0.176
2010	20.752	33.624	-0.235	-0.795	-0.543	-0.231	0.030	0.290
2011	1.494	3.365	-0.043	-0.207	-0.146	-0.048	0.053	0.132
2012	1.156	2.352	0.026	-0.129	-0.066	0.022	0.114	0.185
2013	1.061	2.330	0.117	0.008	0.064	0.121	0.173	0.221

[1] Note that S_ρ and Theil are multiplied by 100.

Table 11: Stochastic Dominance Tests between the Mixed Distributions of Market and Non-Market Values

Year	SD	d	s	Year	SD	d	s	Year	SD	d	s
1976	FSD	**-6.28**	**-6.28**	1989	SSD	51.35	**-12**	2002	No	20.87	117.13
1977	FSD	**-8.16**	**-8.16**	1990	No	44.54	349.9	2003	No	22.37	135.35
1978	No	42.3	32.05	1991	SSD	45.4	**-7.72**	2004	No	26.67	170.83
1979	No	29.4	87.12	1992	No	41.56	293.46	2005	No	17.42	96.57
1980	No	32.54	165.84	1993	SSD	13.73	**-9.96**	2006	No	15.53	71.26
1981	No	33.98	163.2	1994	SSD	19.23	**-8.07**	2007	No	12.22	39.72
1982	SSD	27.29	**-7.41**	1995	SSD	34.21	**-12.94**	2008	No	21.32	153.62
1983	FSD	**-10.84**	**-12.21**	1996	No	16.6	97.13	2009	No	90.25	210.65
1984	FSD	**-10.19**	**-12.87**	1997	No	29.7	213.43	2010	SSD	222.35	**-9.03**
1985	No	29.69	155.07	1998	SSD	48.08	**-10.24**	2011	SSD	65.81	**-8.62**
1986	SSD	41.09	-9.72	1999	No	57.05	437.26	2012	No	71.7	448.04
1987	No	16.39	109.89	2000	No	24.68	178.18	2013	No	10.34	42.31
1988	No	47.88	135.36	2001	No	38.85	290.91				

[1] **No** denotes no (first- or second-order) dominance found, while **FSD** denotes First-order dominance **SSD** denotes Second-order dominance. Only the d and s statistics and corresponding p-values are reported. During the period between 1976-1983, all women dominate working women in either first- or second-order senses. The opposite is observed for all years after 1994, except for 2001 and 2010. These interpretations are based on $d_{1,max}$, $d_{2,max}$, $s_{1,max}$, $s_{2,max}$, which are available in the supplemental material.

(a) Entropy Measures

(b) Conventional Measures

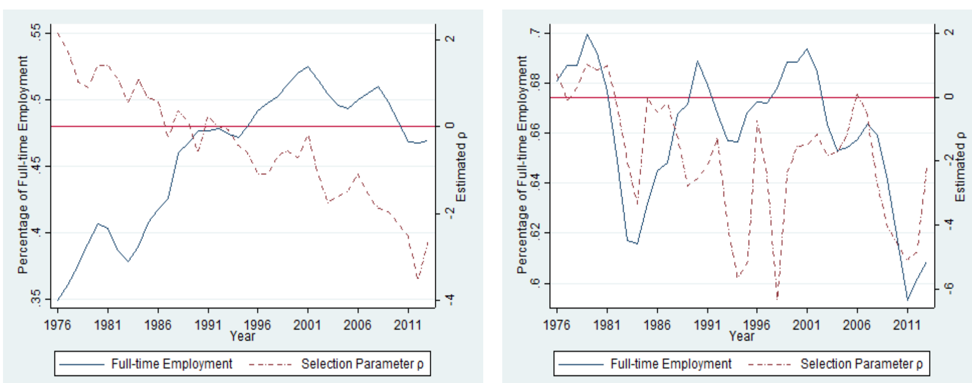
Figure 6: Comparison of Smoothed Trend of The Gender Gap using Mixed Distributions (excluding 2010)

7. Implications

In this section, we use the estimated wage distributions for both men and women to further assess a variety of related assumptions, hypotheses, and findings in the literature on the gender gap and inequality. These investigations are facilitated by simultaneously addressing selection, heterogeneity in outcomes, and the gender gap at the distributional level.

7.1. Selection Bias, Full-time Employment Rates, and Gender Gap

Some of the existing literature has suggested that selection bias may decrease as female employment rates increase. This idea also underlies some previous studies that consider selection as an explanation for the observed relationship between wage- and employment- gaps between men and women across countries. For example, Olivetti and Petrongolo (2008) find that countries with greater gender gap in employment rates (featuring lower female employment rates) are associated with smaller gender gaps. They argue that this is because working women generally have higher wages (i.e., positive selection into labor force), and as more women are employed, selection bias becomes smaller, but the gap increases. Smith and Ward (1989) similarly suggest that the selection bias had become smaller during the 1980s as more women entered the labor force. As pointed out in Mulligan and Rubinstein (2008), such argument is based on the assumption of a fixed, positive selection. Moreover, in the presence of varying selection, the relationship between employment and wage gaps may differ, and the role of selection in explaining the relationship may also be reduced.



(a) Female (b) Male

Figure 7: Percentage of Full-time Female Workers and **Absolute Value** of the Estimated ρ (measuring selection)

Given the estimated selection parameter, ρ, we can first get a sense of how the direction and magnitude of the selection varies over time with employment rates. We plot this relationship in Figure (7). The solid line is the full-time employment rates among women, and the dashed line is the *absolute value* of the estimated ρ. For women, there is negative selection when employment rates are extremely low, while there is positive selection with relatively higher employment. Interestingly, the magnitude of the selection bias becomes smaller as employment rates rise, although the relationship is not necessarily monotonic. For example, selection parameter (negative selection) continued to decline with the increase in employment rates before 1992, and became roughly zero when (full-time) employment rate reached roughly 50% in 1992. Afterward, the magnitude of the selection parameter (positive selection) fluctuates closely with the employment rates. For men, the magnitude of positive selection varied closely with the employment rates. The correlation between employment rates and the absolute values of selection parameters is -0.407 for males, and -0.0752 for females (both statistically insignificant). The latter is indicative of lack of systematic relationship between employment rates and selection bias among women, in contrast to common wisdom.

Table 12: Correlation between Employment Gaps and Various Measures of Wage Gaps (Excluding Year 2010)

	1994-2013		1976-2013	
	Observed Gap	Selection-corrected Gap	Observed Gap	Selection-corrected Gap
	(1)	(2)	(3)	(4)
S_ρ	0.874	0.562	0.973	0.911
Theil	0.869	0.426	0.973	0.886
Mean	0.857	-0.004	0.967	0.563
10th	0.613	-0.066	0.935	0.306
25th	0.796	-0.016	0.974	0.466
50th	0.820	0.035	0.955	0.644
75th	0.780	0.081	0.948	0.747
90th	0.478	-0.005	0.946	0.758

We also compute the correlation between employment gaps and wage gaps. The results are reported in Table (12). Columns (1) and (2) use only the data after 1994 where positive selection prevails. We find that in contrast to Olivetti and Petrongolo (2008), the correlation between employment and wage gaps is instead positive when focusing on only U.S. in the presence of varying selection. Taking into account selection can explain away most of the correlation between employment and wage gaps; for some measures, the correlation is drastically reduced and becomes even statistically insignificant. However, when we include the pre-1994 sample where negative selection existed (Columns 3 and 4), taking into account selection plays a much smaller role in explaining the observed correlation and accounts for a much smaller fraction of the observed correlation. This result is in contrast to the cross-country evidence observed in Olivetti and Petrongolo (2008) where positive selection prevails.

7.2. The Assumption Employed in Blundell et al. (2007)

An alternative approach to address potential selection is to obtain bounds on the *true* gender gap. While requiring milder assumptions on the selection process, such approach often produces very wide or uninformative (worst case) bounds. Further restrictions can tighten the bounds. An example is a form of positive selection imposed in Blundell et al. (2007) to identify the *true* gender gap. This assumption can be expressed as "first-order stochastic dominance of nonworkers wages by that of workers" (Blundell et al. 2007, p.327). This assumption in turn implies that the wage distribution of working women should dominate, in a first-order sense, the wage distribution of the whole population.

Our formal test of this assumption is presented in Table (13), and the corresponding comparisons of the CDFs are in the supplemental material. This assumption fails to hold in 19 out of 38 cases. In many of the early years (for instance, all years during the period of 1976-1983), we instead observe the opposite: the wage distribution of all women dominate, in either first- or second-order senses, the wage

Table 13: Tests of Blundell et al. (2007)'s Assumption (Working Females vs All Females)

Year	SD	d	s	Year	SD	d	s	Year	SD	d	s
1976	FSD	**-0.59**	**-0.91**	1989	No	1.12	3.86	2002	FSD	**-0.93**	**-1.41**
1977	SSD	0.64	**-0.85**	1990	FSD	**-0.85**	**-0.97**	2003	FSD	**-0.93**	**-0.93**
1978	SSD	0.90	**-0.62**	1991	No	0.79	1.87	2004	FSD	**-0.91**	**-1.9**
1979	SSD	1.07	**-0.79**	1992	No	0.81	1.81	2005	FSD	**-0.93**	**-1.11**
1980	SSD	1.04	**-0.78**	1993	No	0.86	2.13	2006	FSD	**-0.89**	**-1.22**
1981	SSD	0.97	**-0.87**	1994	FSD	**-0.87**	**-1.17**	2007	FSD	**-0.98**	**-1.16**
1982	SSD	0.98	**-0.87**	1995	FSD	**-0.75**	**-1.05**	2008	FSD	**-0.95**	**-1.3**
1983	No	1.61	7.55	1996	FSD	**-0.77**	**-1.18**	2009	FSD	**-0.99**	**-1.27**
1984	SSD	1.30	**-0.86**	1997	FSD	**-0.77**	**-1.51**	2010	No	14.16	89.06
1985	No	1.54	10.20	1998	FSD	**-0.75**	**-0.75**	2011	FSD	**-0.92**	**-0.92**
1986	No	1.23	6.04	1999	FSD	**-0.84**	**-1.26**	2012	FSD	**-0.95**	**-1.31**
1987	No	0.74	0.84	2000	FSD	**-0.95**	**-1.26**	2013	FSD	**-0.86**	**-1.24**
1988	No	1.27	4.55	2001	FSD	-1.03	-2.05				

[1] **No** denotes no (first- or second-order) dominance found, while **FSD** denotes First-order dominance **SSD** denotes Second-order dominance. Only the d and s statistics and corresponding p-values are reported. During the period between 1976-1983, all women dominate working women in either first- or second-order senses. The opposite is observed for all years after 1994, except for 2001 and 2010. These interpretations are based on $d_{1,max}$, $d_{2,max}$, $s_{1,max}$, $s_{2,max}$, which are available in the supplemental material.

distribution of full-time women. These dominance relations are statistically significant. We observe evidence supporting the assumption only after 1994 (except 2001 and 2010). This result does not necessarily mean that it would fail to hold in the U.K., the country studied in Blundell et al. (2007). Note that women's employment rates are relatively higher in the U.K. than U.S.. They show that the employment rates generally range between 60% and 75%, which is well beyond the region of negative selection as indicated above. However, our results do not support applicability of this assumption in the U.S..

7.3. Within-Group Inequality and Selection Patterns

The result that selection for women varies from negative to positive over time but remains mostly positive for men casts doubts on some hypotheses and assumption used in the literature (Sections 7.1 and 7.2). The question is: What can explain the change of the selection pattern for women? And what can explain the stability in the selection pattern for men?

An explanation is put forth in Mulligan and Rubinstein (2008). Using the traditional Roy model, they argue that increasing wage inequality within gender over time would cause women to invest more in their market productivity and lead abler and hence higher-wage women/men to participate in labor force. In other words, increasing wage inequality within gender is associated with positive selection. This explanation is seemingly supported by the findings in the inequality literature. For instance, influential studies such as Autor et al. (2008) indeed find that the inequality, especially the upper-tail inequality, has increased steadily for both men and women in the past decades.

However, these findings are based on *observed* wage inequality, (without considering individuals who do not work), instead of the underlying wage inequality, upon which Mulligan and Rubinstein (2008)'s explanation is built. These two quantities could be drastically different. Not only could addressing the selection and estimating the distributions of potential wages potentially invalidate Mulligan and Rubinstein (2008)'s explanation. It may also challenge the conventional wisdom about the patterns in the within-group inequality.

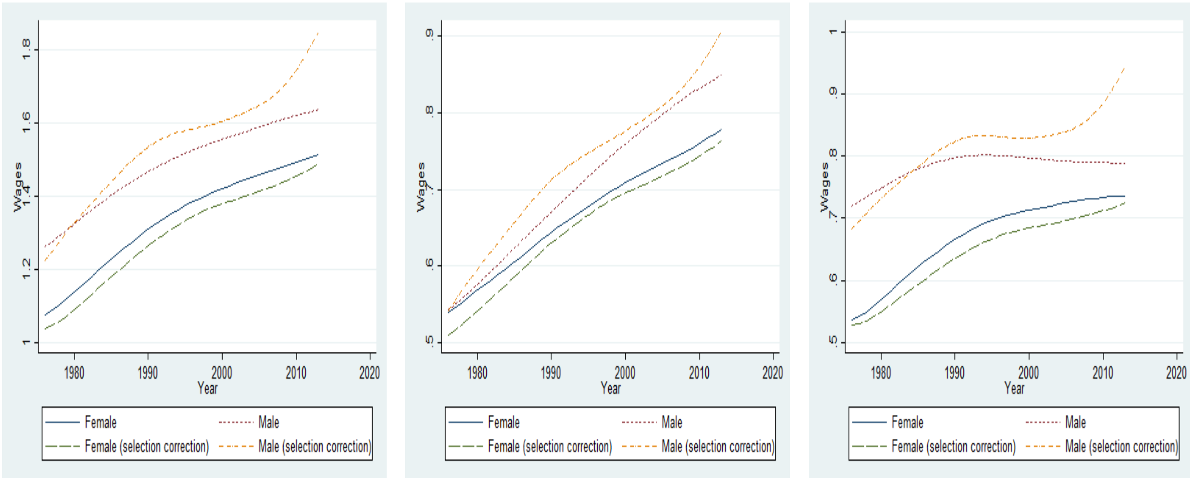
Because our approach recovers the distributions of the potential wage outcomes, this allows us to recover the patterns in the underlying wage inequality, as well as to formally test this explanation and verify the findings in the inequality literature. Here we provide the smoothed time trend of the wage inequality measures used in Mulligan and Rubinstein (2008) to better visualize the evolution of the inequality in Figure (8): the difference of the 90^{th} and 10^{th} percentiles of the log wages. In addition, we also plot the difference of the 90^{th} and 50^{th} percentiles, as well as the difference of the 50^{th} and 10^{th} percentiles.

First, without considering non-working individuals, our findings are broadly consistent with Mulligan and Rubinstein (2008) and Autor et al. (2008). Specifically, the trend of the *observed* within-group inequality for both men and women (without considering non-working individuals) increased rapidly during the period between early 1980s and mid-1990s but at a slower rate afterward. The patterns of the upper-tail (90/50) inequality and lower-tail (50/10) inequality differ by gender. For men there exists a divergence in the time trend of the upper- and lower-tail inequalities. Specifically, the lower tail inequality has exhibited an increasing trend during the 1980s, but the trend stagnated at the same level afterward, while the upper-tail inequality has exhibited a rapidly increasing trend in the past decades. However, the pattern for women is different. Specifically, there has been a steadily increasing trend for *both* upper- *and* lower-tail inequalities since the mid-1970s.[36]

Second, while we find that the inequality measures for women after taking into those who do not work are generally larger than the inequality measures failing to do so, the general patterns of increasing trend for these measures continue to hold. By contrast, while the overall and upper-tail inequality measures (90/10 and 90/50) for men continue to exhibit an increasing trend, the lower tail inequality (50/10) measure also shows an increasing trend that was at an even faster rate during the most recent recession, which is contradictory to the results above. This latter result questions the common finding in the inequality literature that the increase in the overall inequality is only attributed to the increase in the upper tail inequality, but not the lower tail inequality. This common finding is likely to be a result of failure to take into account those men who are not working and likely earn less wages.

Further analysis indicates important differences between our results and those in Mulligan and Rubinstein (2008) and Autor et al. (2008), even for those similar ones. These differences stem from the sources of the increased inequality. To see this, we plot the smoothed trend of select percentiles for both men and women (before and after the selection is controlled for) in Figure (9). It can be seen that the reason for the increase in the *observed* overall inequality (in their paper) is due to the drastic increase in the 90^{th} percentile of the distribution (without selection correction), while the 10^{th} percentile remained relatively stable and exhibited some increasing trend during this period. By

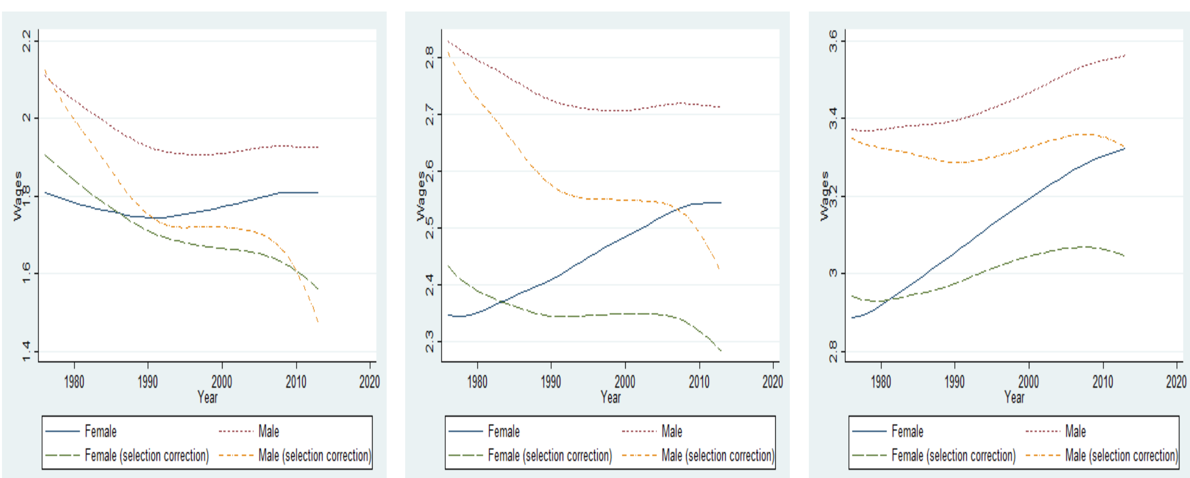
[36] Note that Autor et al. (2008) use the CPS data that begin at 1963.



(a) 90-10 (b) 90-50 (c) 50-10

Figure 8: Comparison of Smoothed Trend of Various Inequality Measures

contrast, using the true population wage distribution shows that the increase in wage inequality among women is actually a combined result of the decline at the 10^{th} percentile *and* an increase at the 90^{th} percentile of the distribution with selection correction. The change in the extreme lower tail is missed in the standard analysis since women who receive lower wage offers choose not to work, and the wages in the lower tail are therefore "inflated". The increasing trend for the 90^{th} percentile of women's wage distribution is similarly over-estimated when failing to take into account non-full-time employed women. Note also that the pattern for women after the selection is addressed is surprisingly close to the one for men, where we find a similar divergence in wages between the most skilled (90^{th} percentile) and the least skilled (10^{th} percentile) men. These results are also broadly consistent with and lend further support to the inequality literature on this issue (e.g., Juhn et al. 1993).



(a) 10^{th} (b) 50^{th} (c) 90^{th}

Figure 9: Comparison of Smoothed Trend of The Distributions for both Men and Women with and without Selection Correction

In sum, the implied evolution of wage inequality within and between gender is consistent with the proposed theoretical explanation in Mulligan and Rubinstein (2008), while the patterns for men and the sources of the patterns are different from the inequality literature. There may be alternative explanations, for example, differential impacts of expanded child care availability on women, but we focus on the one that can be directly tested and related to the gender gap and inequality literature.[37]

7.4. Decomposition and Counterfactual Analysis of the Gender Gap in the presence of Selection

Decomposition/counterfactual analysis is informative about potential sources of the gender gap. Such analysis constructs the counterfactual wage distribution when either wage "structure" or the distribution of human capital characteristics ("composition effects") for women is varied, holding the other fixed. Comparison with counterfactual distributions provides a decomposition of the gender gap between the two components. Such analysis has a long-standing history in labor economics (see Altonji and Blank (1999) and Fortin et al. (2011) for excellent accounts of this issue). However, most of this type of analysis usually ignores the selection issue, and is at the mean. We share a view with Aaberge et al. (2013) and Carneiro et al. (2001) that there is a need to go beyond decomposing the gender gap and simply obtaining the counterfactual effects at select quantiles. In this section, we illustrate how addressing these issues may impact the standard counterfactual or decomposition analysis.

Structural effects are objects of policies promoting equal wage-setting; for example, equity programs that are designed to address wage differences between men and women with the same skills and work by equalizing their pay structures. The composition effects concern human capital characteristics such as education. Policy/treatment outcomes may produce "winners" and "losers"; structural (or composition) effects could be positive at some parts of the distributions while negative at others. Once the counterfactual distributions (with and without correction for selection) are obtained, our metric entropy gap and SD analysis can be employed.

As noted earlier, counterfactual distributions may be based on conditional quantile regressions, or on re-weighting by propensity scores. We adopt the first approach because we are able to estimate the (*true*) conditional quantile selection regressions.[39]

Machado and Mata (2005) is among the first to estimate quantiles to recover the counterfactual

[37] Another possible explanation is differential impacts of expanded child care availability on women. The impact of expanded availability of child care over time on women's employment may vary with women's income levels. According to census data, child care options outside of the home have increased drastically in the past decades. Specifically, the number of child care facilities increased from $262,511$ in 1987 to $766,401$ in 2007, a threefold increase.[38] Meanwhile, the child care costs have also been increasing but affected families at different income levels very differently. According to the census data in 2011, families with employed mothers whose monthly income was $4,500$ or more paid roughly 6.7% of their family income (an average of 163 a week) for child care, while those with monthly income of less than $1,500$ paid about 40% of their family income (an average of 97 a week). Source: `http://www.pewresearch.org/fact-tank/2014/04/08/rising-cost-of-child-care-may-help-explain-increase-in-stay-at-home-moms/` Moreover, child care subsidies for low-income women "remain inadequate" (Blank 2006). As a result, the low-wage women may not be able to afford child care by working, while high-skilled women, on the other hand, could continue their careers because they could afford it. This implies that over time it is more likely to observe high-wage earners enter the full-time employment while low-wage earners struggle to juggle work and family. There indeed exists evidence that daycare/pre-school participation rates are higher for high-income families (Landerso and Heckman, 2016).

[39] The re-weighting approach cannot be readily extended to address selection for decomposition for the *whole* population. In a companion paper (Maasoumi and Wang, 2016), which examines the racial gap among women, we extend the results in Huber (2014) and propose a re-weighting approach based on *nested* propensity score to recover the counterfactual distributions for the *selected* population.

distribution, and Chernozhukov et al. (2013) discuss the corresponding inferential theory.[40] The counterfactual distribution can be recovered as follows,

$$F_{Y_c}(y) = F_{Y\langle i|j\rangle}(y) = \int F_{Y_i|X_i}(y|x)dF_{X_j}(x) \tag{18}$$

From Equation (13), it follows that Equation (18) can be re-written as follows

$$\begin{aligned} F_{Y_c}(y) = F_{Y\langle i|j\rangle}(y) &= \int\{\int_0^1 I[Q_\tau(Y_i|X_i) \leq y]d\tau\}dF_{X_j}(x) &(19) \\ &= \int\{\int_0^1 I[X\beta_i \leq y]d\tau\}dF_{X_j}(x) &(20) \end{aligned}$$

The last equality follows from our specification of the conditional quantile model. We can identify the following counterfactual outcome distributions:

$$\begin{aligned} F_{Y_{c1}}(y) &= \int\{\int_0^1 I[X\beta_m \leq y]d\tau\}dF_{X_f}(x) \quad \text{(Counterfactual Distribution \#1)} &(21) \\ F_{Y_{c2}}(y) &= \int\{\int_0^1 I[X\beta_f \leq y]d\tau\}dF_{X_m}(x) \quad \text{(Counterfactual Distribution \#2)} &(22) \end{aligned}$$

F_{c1} represents the counterfactual distribution when male wage structure is used, holding the distribution of women's human capital characteristics unchanged. F_{c2} represents the counterfactual distribution when female wage structure is used, holding the distribution of men's human capital characteristics unchanged. The differences in the distributions F_{c1} and F_1 provide insight into "structural effects". The differences in F_{c2} and F_1 come from differences in the distribution of human capital characteristics; the "composition effects".

Various measure of the gap are presented in Tables (A.17)-(A.20) in the supplemental material. We summarize the findings here. First, regardless of whether we control for selection, structural effect appears to be much greater than the composition effect. The latter is rather small and often close to zero. Second, failure to control for selection often underestimates the role of composition effects in contributing to the gender gap. When not considering selection, we find that except in few early years, changing the distribution of characteristics would not improve women's wages; instead it could hurt their labor market outcomes (evident from the negative distance at select percentiles). This negative effect started as early as 1980 for women in the upper tail of the wage distribution. This result seems to suggest that women not only have caught up with men but could also have surpassed them in the level of human capital characteristics (captured by our covariates). This result is consistent with the increasingly widening gap in college education between men and women. For example, Goldin et al. (2006) find that "by 1980, the college gender gap in enrollments had evaporated" and call this change a "homecoming" of American college women (to the parity observed in the early twentieth century). However, when controlling for selection, we find that the effects of such progress on the gender gap

[40] Albrecht et al. (2009) extends this framework to address the selection issue at the distributional level. However, Albrecht et al. (2009)'s approach is based on Buchinsky (2001)'s quantile selection model, which, as argued above, relies on a rather restrictive wage structure.

are overestimated. Specifically, changing the distribution of the observed characteristics could still be beneficial and improve women's wage outcomes for those in the lower tail before 1994 (nearly a half of the time span that we examine), and for those in the upper tail until the beginning of the twenty-first century. Note that for composition effects, the difference between the selection-corrected and uncorrected stem from not only the differences in human capital composition, but also the differences in the bias in estimation of wage structures. Finally, regardless of whether we control for selection, structural effects are positive and substantial, suggesting that changing women's pay structure could be beneficial (and implying that discrimination may exist). Addressing the selection impacts the estimates of such beneficial effects (or the severity of discrimination). Failure to address the selection could generally overestimate the structural effects for women in the lower tail of the wage distribution (especially before the twenty-first century), while it generally underestimates the structural effects for women in the upper tail (especially since mid-90s).

Turning to SD tests in Tables (A.22)-(A.24) in the supplemental material, when not controlling for selection, the female counterfactual distributions with male wage structure first-order dominate the female wage distribution. This result implies that, in these cases, changing earnings structure would result in a change in the earnings distribution for women, and that the change is *uniformly* in favor of all women. Taken at their face values, such results are qualitatively consistent with the prior findings that such policies as equity pay could be potentially successful in closing the gender gap (e.g. Hartmann and Aaronson, 1994; Gunderson and Riddell, 1992). These results are even stronger than what is implied by various measures of the gap above. However, such strong results do not necessarily hold when selection is addressed. Once selection is controlled for, we fail to find dominance relations in 10 out of 38 cases, and such results arise because women in the extreme lower tail do not necessarily benefit from change in wage structure, an important result masked by examining only the gap at select percentiles.

When we do not control for selection, second-order dominance is inferred in early years only, but no meaningful SD ranking of the female wage distribution and the counterfactual wage distribution (#2) in later years. This result implies that even with aversion to inequality, changing the distribution of women's human capital characteristics to the distribution of men's characteristics may not represent welfare improvement. As discussed above, an implication is that the *distribution* of women's human capital characteristics is not necessarily inferior to that of men's, and thus policies aimed at changing the human capital characteristics only, may not produce relative improvements for women. However, once selection is controlled for, we find that the results are drastically different. We instead observe first-order dominance relations in every year.

These results indicate potentially misleading policy conclusions by failure to account for selection, especially in regards to composition effects. [41]

[41] Certain assumptions underly this type of analysis deserve closer examination. As noted in Fortin et al. (2011), one standard assumption is that of conditional independence. This may fail to hold if a variable is endogenous and correlated with the unobservables (e.g., cognitive and non-cognitive skills; see Heckman et al. (2006)). Some recent literature also suggests that psychological and socio-psychological factors (e.g., risk preferences) may help to explain the gender gap. However, as noted in Bertrand (2010) notes that such information is largely limited to the laboratory setting (not in a large-scale data like CPS); and that the existing research in this areas "is clearly just in its infancy and far from conclusive, with many contradictory findings." However, as pointed out in Fortin et al. (2011), while we cannot identify

8. Summary of Main Contributions, Findings and Conclusions

This paper examines two issues in measuring the overall gender gap in U.S., namely heterogeneity in wages and selection into full-time employment. In the case of heterogeneity, we find that aggregation of the gap at all quantiles is helpful as a summary measure. Selection is a significant issue, as is heterogeneity. The entropy gap, uncorrected for selection, indicates greater convergence of women and men's earnings in the early years but much slower convergence afterward, compared to those found in the literature. Stochastic dominance rankings provide robustification. Wage distribution for men first order dominate women's. This conclusion is robust to a wide class of increasing Evaluative Functions (EFs). Without selection, *any* measure would be adequate for "ordering" outcomes, but would differ in magnitude.

Once selection is accounted for, no dominance relation holds in quite a few cases in early years. Only narrower preference functions that are more than merely increasing and concave would rank men wages over women's. Our entropy measures suggest that there was a much slower decline in the gender gap when selection is corrected. There was even a reversal in the declining trend over portions of the distribution in the years between mid-1990s and the most recent recession, which is missed in the baseline results. During the most recent recession, there was a clearer declining trend in the gap among low-skilled workers. A similar pattern is observed for some groups by education and race. We find that women with the least education or black women appear to have witnessed much smaller progress in catching up with their male counterparts, especially during recent years.

A new alternative gender gap is proposed (based on the mixed distribution of wage offers and reservation wages conditional on employment status). It implies that women's relative well-being may have deteriorated over time, again in contrast to the baseline results.

We further revisit and challenge some important findings, hypotheses, and assumptions in the literature on the gender gap and inequality. In contrast to the cross-country findings in Olivetti and Petrongolo (2008), we find that in the presence of varying selection, selection plays a much smaller role in explaining the observed correlation between employment and the gender wage gap in the U.S.. Using the estimated wage functions and distributions, we formally test the assumption of first-order dominance relation, which is assumed in Blundell et al. (2007), and reject it. Furthermore, we find that there exists an increasing trend of within-group inequality in *both* the upper *and* lower tail of the distributions among both men and women. This result provides empirical support for the theoretical explanation proposed in Mulligan and Rubinstein (2008) to explain the pattern of varying selection for women over time, from negative to positive. It also challenges the conventional wisdom that the increased overall inequality among men is only attributed to the increase in the upper tail, but not the

the contribution of education vs ability in this context, the aggregate decomposition nevertheless remains valid provided that ignorability holds. For example, even though we may expect unobserved ability to be correlated with education, it is reasonable to assume that there exist no systematic differences between men's and women's innate ability, given education and their characteristics. In that case, the aggregate decomposition remains valid. The work of attribution of wages to various covariates is not a focus of this particular paper, which addresses the question of "what is the gap" and related welfare implications, and thus is left for future research. Note that the quantile based joint distribution approach can potentially address the failure of CIA. But this will require an IV for each endogenous variable in the wage equation. If that challenge were to be successfully met, one would use IV quantile regressions (e.g., Chernozhukov and Hansen 2008). This approach may become infeasible given the number of variables that are typically included in the wage equations.

lower tail, of wage distribution. While the work of attribution of the gap to separate covariates and sources remains a major undertaking, our preliminary decomposition analysis suggests that differences in wage structures may likely be the main cause of the differences in wages between men and women.

Our approach can be extended to multidimensional gender gap, which has not been rigorously studied before. As is commonly acknowledged, well-being is in general a multi-dimensional concept, and earnings is potentially only "a vague reflection of societal wellness" (Anderson et al. 2014); (Sen 1992, p.46). Our approach could be particularly useful because both entropy measures and SD analysis are constructed over the space of *distributions* and can be seamlessly applied to univariate and multi-outcome contexts.

References

Aaberge, R., T. Havnes, and M. Mogstad. 2013. "A Theory for Ranking Distribution Functions." *Unpublished Manuscript* .

Albrecht, James, Anders Björklund, and Susan Vroman. 2003. "Is There a Glass Ceiling in Sweden?" *Journal of Labor Economics* 21:145–177.

Albrecht, James, Aico van Vuuren, and Susan Vroman. 2009. "Counterfactual distributions with sample selection adjustments: Econometric theory and an application to the Netherlands." *Labour Economics* 16:383 – 396.

Altonji, J.G. and R.M. Blank. 1999. *Handbook of Labor Economics*, volume 3, chapter Race and Gender in the Labor Market, pp. 3143–3259. New York: Elsevier.

Anderson, G., M.G. Pittau A. Farcomeni, and R. Zelli. 2014. "A New Approach to Measuring and Studying the Characterstics of Class Membership: The Progress of Poverty, Inequality and Polarization of Income Classes in Urban China." *Unpublished Manuscript* .

Angrist, J.D., V. Chernozhukov, and I. Fernandez-Val. 2006. "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure." *Econometrica* 74:539–563.

Angrist, J.D. and J. Pischke. 2009. *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.

Arellano, M. and S. Bonhomme. 2016. "Sample Selection in Quantile Regression: A Survey." *Handbook of Quantile Regression* .

Arellano, M. and S. Bonhomme. forthcoming. "Quantile Selection Models: with an Application to Understanding Changes in Wage Inequality." *Econometrica* .

Arrar, A. and J. Duclos. 2003. "An Atkinson-Gini Family of Social Evauation Functions." *Economics Bulletin* 3:1–9.

Atkinson, A. 1970. "On the Measurement of Inequality." *Journal of Economic Theory* II:244–63.

Autor, D.H., L.F. Katz, and M.S. Kearney. 2008. "Trends in U.S. Wage Inequality: Reivsing the Revisionists." *Review of Economics and Statistics* 90:300–323.

Basu, S., J.G. Fernald, and M.D. Shapiro. 2001. "Productivity Growth in the 1990s Technology, Utilization, or Adjustment?" *Carnegie-Rochester Conference Series on Public Policy* 55:117–65.

Bertrand, M. 2010. *Handbook of Labor Economics*, volume 4B, chapter New Perspective on Gender, pp. 1545–1592. New York: Elsevier.

Blank, R.M. 2006. "Was Welfare Reform Successful?" *Economists' Voice* 3:Article 2.

Blau, F.D. 1998. "Trends in the Well-Being of American Women, 1970-1995." *Journal of Economic Literature* XXXVI:112–165.

Blau, F.D. and L.M. Kahn. 1994. "Rising Wage Inequality and the U.S. Gender Gap." *American Economic Review* 84:23–28.

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

Blau, F.D. and L.M. Kahn. 2006. "The U.S. Gender Pay Gap in the 1990s: Slowing Convergence." *Industrial and Labor Relations Review* 60:45–66.

Blau, F.D. and L.M. Kahn. 2013. "Female Labor Supply: Why is the US Falling Behind?" *IZA DP* 7140.

Blau, F.D. and L.M. Kahn. 2016. "The Gender Wage Gap: Extent, Trends, and Explanations." *NBER Working Papers* .

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir. 2007. "Changes in the Distribution of Male and Female Wages Accounting For Employment Composition Using Bounds." *Econometrica* 75:323–363.

Bollinger, Christopher R. and Barry T. Hirsch. 2006. "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching." *Journal of Labor Economics* 24:483–519.

Bollinger, Christopher R. and Barry T. Hirsch. 2013. "Is Earnings Nonresponse Ignorable?" *The Review of Economics and Statistics* 95:407–416.

Bonhomme, S., G. Jolivet, and E. Leuven. 2014. "School Characteristics and Teacher Turnover: Assessing the Role of Preferences and Opportunities." *Unpublished Manuscript* .

Buchinsky, M. 1998. "The Dynamics of Changes in the Female Wage Distribution in the U.S.A.: A Quantile Regression Approach." *Journal of Applied Econometrics* 13:1–30.

Buchinsky, M. 2001. "Quantile Regression with Sample Selection: Estimating Women's Return to Education in the U.S." *Empirical Economics* 26:87–113.

Card, D. and J.E. DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics* 20:733–783.

Carneiro, P., K.T. Hansen, and J.J. Heckman. 2001. "Removing the Veil of Ignorance in Assessing the Distributional Impacts of Social Policies." *Swedish Economic Policy Review* 8:273–301.

Chamberlain, G. 1986. "Asymptotic Efficiency in Semi-parametric Models with Censoring." *Journal of Econometrics* 32:189–218.

Chang, S.K. 2011. "Simulation estimation of two-tiered dynamic panel Tobit models with an application to the labor supply of married women." *Journal of Applied Econometrics* 26:854–871.

Chernozhukov, V., I. Fernandez-Val, and B. Melly. 2013. "Inference on Counterfactual Distributions." *Econometrica* 81:2205–2268.

Chernozhukov, V. and C. Hansen. 2008. "Instrumental variable quantile regression: A robust inference approach." *Journal of Econometrics* 142:379–298.

Cohn, D., G. Livingston, and W. Wang. 2014. "After Decades of Decline, A Rise in Stay-at-Home Mothers." Research report, Pew Research Center.

Council of Economic Advisers. 2016. "The Long-Term Decline in Prime-Age Male Labor Force Participation." Technical report, Executive Office of the President of the United States, Washington DC.

DiNardo, J., N. Fortin, and T. Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica* 64:1001–1044.

Doksum, K. 1974. "Empirical probability plots and statistical inference for nonlinear models in the two-sample case." *Annals of Statistics* 2:267–77.

Firpo, S. 2007. "Efficient Semiparametric Estimation of Quantile Treatment Effects." *Econometrica* 75:259–276.

Fortin, N., T. Lemieux, and S. Firpo. 2011. *Handbook of Labor Economics*, volume 4, chapter Decomposition Methods in Economics, pp. 1–102. Elsevier.

Frolich, M. 2007. "Propensity Score Matching Without Conditional Independence Assumption – With An Application to the Gender Wage Gap in the United Kingdom." *Econometrics Journal* 10:359–407.

Goldin, C. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104:1091–1119.

Goldin, C., L.F. Katz, and I. Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives* 20:133–156.

Granger, C., E. Maasoumi, and J.C. Racine. 2004. "A dependence metric for possibly nonlinear processes." *Journal of Time Series Analysis* 25:649–669.

Gunderson, M. and W.C. Riddell. 1992. "Comparable Worth: Canada's Experience." *Contemporary Policy Issues* 10:85–94.

Hartmann, H.I. and S. Aaronson. 1994. "Pay Equity and Women's Wage Increases: Success in the States, A Model for the Nation." *Duke Journal of Gender Law and Policy* 1:69–87.

Heckman, J.J. 1974. "Shadow Price, Market Wages, and Labor Supply." *Econometrica* 42:679–694.

Heckman, J.J. 1980. *Female Labor Supply: Theory and Estimation*, chapter Sample Selection Bias as a Specification Error with an Application to the Estimation of Labor Supply Functions. NJ: Princeton University Press.

Heckman, J.J. 1990. "Varieties of Selection Bias." *American Economic Review* 80:313–318.

Heckman, J.J., J. Stixrud, and S. Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 3:411–482.

Hirsch, B.T. and E.J. Schumacher. 2004. "Match Bias in Wage Gap Estimates Due to Earnings Imputation." *Journal of Labor Economics* 22:689–722.

Huber, M. 2014. "Treatment Evaluation in the Presence of Sample Selection." *Econometric Reviews* 33:869–905.

Huber, M. and G. Mellace. 2014. "Testing Exclusion Restrictions and Additive Separability in Sample Selection Models." *Empirical Economics* 2011-45.

Juhn, C. and K.M. Murphy. 1997. "Wage Inequality and Family Labor Supply." *Journal of Labor Economics* 15:72–97.

Juhn, C., K.M. Murphy, and B. Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy* 101:410–442.

Keane, M.P., K.I. Wolpin, and P. Todd. 2011. *Handbook of Labor Economics*, volume 4, chapter The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications, pp. 331–461. Elsevier.

Koenker, R. and G. Bassett. 1978. "Regression Quantiles." *Econometrica* 46:33–50.

Kolm, S.C. 1969. *Public Economics*, chapter The Optimal Production of Social Justice, pp. 145–200. London: MacMillan.

Landerso, R. and J.J. Heckman. 2016. "The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the U.S." *IZA Disussion Paper* 10000.

Lemieux, Thomas. 2006. "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *The American Economic Review* 96:461–498.

Li, Q. and J.S. Racine. 2008. "Nonparametric Estimation of Conditional CDF and Quantile Functions With Mixed Categorical and Continuous Data." *Journal of Business & Economic Statistics* 26:423–434.

Linton, O., E. Maasoumi, and Y.J. Whang. 2005. "Consistent Testing for Stochastic Dominance: A Subsampling Approach." *Review of Economic Studies* 72:735–765.

Maasoumi, E. 1993. "A Compendium to Information Theory in Economics and Econometrics." *Econometric Reviews* 12:1–49.

Maasoumi, E. 1998. *The Handbook of Applied Econometrics, Vol II: Microeconomics*, chapter Empirical Analyses of Inequality and Welfare. Hoboken, NJ: Wiley Blackwell Publishers.

Maasoumi, E. 2001. *Companion to Econometric Theory*, chapter Parametric and Nonparametric Tests of Limited Domain and Ordered Hypotheses in Economics. Hoboken, NJ: Wiley Blackwell Publishers.

Maasoumi, E. and J.C. Racine. 2002. "Entropy and predictability of stock market returns." *Econometric Reviews* 107:291–312.

Maasoumi, E. and L. Wang. 2016. "What Can We Learn about the Racial Gap in the Presence of Sample Selection." *Working paper* .

Machado, C. 2012. "Selection, Heterogeneity and the Gender Wage Gap." *IZA Discussion Papers* 7005.

Machado, J. and J. Mata. 2005. "Counterfactual Decompositions of Changes in Wage Distributions Using Quantile Regression." *Journal of Applied Econometrics* 20:445–65.

Martins, M. 2001. "Parametric and Semiparametric Estimation of Sample Selection Models: An Empirical Application to the Female Labour Force in Portugal." *Journal of Applied Econometrics* 16:23–39.

Meester, S. and J. MacKay. 1994. "A Parametric Model for Cluster Correlated Categorical Data." *Biometrics* 50:954–963.

Melly, B. 2005. "Decomposition of Differences in Distribution Using Quantile Regression." *Labour Economics* 12:577–590.

Melly, B. and M. Huber. 2008. "Sample Selection, Heteroskedasticity, and Quantile Regression." *Unpublished Manuscript* .

Melly, B. and M. Huber. 2011. "Nonseparable Sample Selection Models." *Unpublished Manuscript* .

Mora, R. 2008. "A Nonparametric Decomposition of the Mexican American Average Wage Gap." *Journal of Applied Econometrics* 23:463–485.

Mulligan, C.B. and Y. Rubinstein. 2008. "Selection, Investment, and Women's Relative Wages Over Time." *Quarterly Journal of Economics* 123:1061–1110.

Neal, D.A. 2004. "The Measured Black-White Wage Gap among Women Is Too Small." *Journal of Political Economy* S1:S1–S28.

Nopo, H. 2008. "Matching as a Tool to Decomposte Wage Gaps." *Review of Economics and Statistics* 90:290–299.

Oaxaca, R.L. 1973. "Male/Female Wage Differentials in Urban Labor Markets." *International E-conomoic Review* 14:693–709.

Oaxaca, R.L. and M.R. Ransom. 1994. "On the Discrimination and Decomposition of Wage Differentials." *Journal of Econometrics* 61:5–21.

Olivetti, C. and B. Petrongolo. 2008. "Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps." *Journal of Labor Economics* 26:621–654.

Pratt, J.W. 1964. "Risk Aversion in the Small and in the Large." *Econometrica* 32:122–136.

Sahin, A., J. Song, and B. Hobijn. 2010. "The Unemployment Gender Gap During the 2007 Recession." *Federal Reserve Bank of New York Current Issues* 16:1–7.

Sen, A. 1992. *Inequality Reexamined*. Cambridge: Harvard University Press.

Smith, J.P. and M.P. Ward. 1989. "Women in the Labor Market and in the Family." *Journal of Economic Perspectives* 3:9–23.

StataCorp. 2009. *Stata 11 Base Reference Manual*. Stata Press, College Station, TX.

Trivedi, P.K. and D.M. Zimmer. 2005. "Copula Modeling: An Introduction for Practitioners." *Foundations and Trends in Econometrics* 1:1–111.

Waldfogel, J. and S.E. Mayer. 2000. *Finding Jobs: Work and Welfare Reform*, chapter Gender Differences in the Low-Wage Labor Market, pp. 193–232. New York, New York: Russell Sage Foundation.