The Gender Gap Between Earnings Distributions

Esfandiar Maasoumi
Emory University

Le Wang
University of Oklahoma, & IZA

Abstract

We advocate a different approach to measure the gender gap, which summarizes each distribution by suitable evaluative functions, and then computes the difference between the evaluations. This approach does not assume rank invariance as the conventional approach. We discuss the decision-theoretic framework behind different functions, and introduce a class of measures based on entropy functions. We further adopt quantile-Copula approaches to account for selection into full-time employment, and also discuss how we can take into account non-market values in measuring the gap. The evolution of the gender gap depends on the measure of it, and whether non-market values are incorporated. We further assess and challenge a variety of assumptions, hypotheses, and findings in the literature.

JEL classifications: J1, J3, J7, C21, D6, I3

Keywords: Gender gap, selection, quantiles, copula, entropy, counterfactual distributions, treatment effect, decision theory, inequality, Hellinger, value of time, labor market, employment, wages

—

† We specially thank Professor James Heckman (the Editor) for very generous and extensive comments and guidance, and six knowledgeable referees for their constructive comments and suggestions, and Stephane Bonhomme for inspiring discussions of the findings. Part of the work 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 Karen Conway, Sol Polachek, Chunbei Wang, and seminar participants at numerous universities and conferences for their helpful comments.

1. Introduction

Measuring the gender wage gap is vital for our understanding of women's well-being relative to men's in society. We contribute to this vast literature in two main ways. The first contribution of the paper is to expose the conceptual problems and implicit assumptions in the conventional approach to measure the gender gap, and then to provide a new approach that addresses them. The second contribution is to understand how important is accounting for selective participation and its trends in *any* measure of the gender gap.

Measuring the gender gap is about comparing two wage distributions, and there can be two approaches. One approach is to compute quantile-by-quantile gaps and summarize these by suitable evaluative functions. A second approach is to first summarize (characterize) each distribution by suitable evaluative functions, and then compute the difference between the evaluations. The popular approach is the first approach, but we advocate the second (for the first time in the literature on the gender gap). Sometimes these two approaches coincide, such as for the gap at the mean (average treatment effect).[1]

The difference between the two approaches is subtle. The first approach relies on identification of the gap at quantiles. To see the problems of this approach, consider a society with only two men (Males A and B), and two women (Females A and B). Male wages are ($5,000 , $1,200), respectively; Female wages are ($3,000 , $1,000). Quantile-by-quantile analysis will compare Female A to Male A, and Female B to Male B. However, occupying the same rank in their respective group does not necessarily mean that Female A and Male A are comparable individuals. Implicitly assumed and required in the quantile comparisons is an assumption of rank invariance (or similarity), i.e., one's relative rank is preserved when endowed with each other's skill sets or market returns. Rank invariance requires that male and female ranks refer to the same skills and substitutions thereof, or at least the same intrinsic values of skills.

Rank invariance is unlikely to be satisfied empirically, and we indeed reject this assumption for several decades of CPS data in the US. Without rank invariance, it is questionable that the first approach ("quantile treatment effect") can deliver meaningful measures of the gender gap. Our approach (to summarize the distribution first and then compare the summary measures) addresses this issue because it is concerned with the distributions, instead of individuals. Our approach can deliver a measure of the gender gap consistent with our goal of understanding women's well-being, as a group (and its subgroups) relative to men's, also *as a group*.

There is no universally accepted evaluation function of a distribution, and there are many candidates for our proposed approach. Averages, inequality measures and entropies are all well-known functions of distributions that summarize its quantiles *anonymously*, without regard to identity of those who occupy a given quantile. Each function attributes its own weights to different wage levels. For example, the average (function) assumes equal weights to all percentiles, treating a dollar of high-wage earners and a dollar of low-wage earners equally. Evaluative functions that underly various gender gap measures have a decision theoretic basis, of which we provide a brief account in Section 2.

The decision theoretic framework enables us to discuss alternative evaluation criteria of the wage distributions other than mean, and the corresponding measure of the gender gap. Specifically, we introduce a flexible family of measures of the gender gap based on entropy functions (the Generalized Entropy family

[1]The mean gap can be thought of as averaging the quantile gaps, or taking the difference in averages of each distribution.

that includes a normalized Kullback-Leibler-Theil measure, and a normalized Hellinger measure). Entropy functions share similarities to characteristic functions, such as a one-to-one relation to the corresponding distribution. More importantly, unlike the average function, entropies satisfy many desirable properties such as aversion to inequality (which assigns more weights to a dollar of transfer at lower wages than at higher wages; i.e, the Pigou-Dalton principle of transfers). Each entropy function in this class is characterized by a different level of inequality aversion (for an impartial observer/evaluator of the wage distributions).[2]

It is worth emphasizing that what we advocate is the use of the decision theoretic framework to evaluate wage distributions, but not a specific evaluation criteria or measure. The non-uniqueness of evaluation functions indeed leads us to further provide statistical tests for stochastic dominance rankings, which identifies situations in which one wage distribution may be preferred to another, irrespective of specific preferences (or weights to different wage levels) within a given class. Both rank invariance and choice of evaluation functions are problems that have not received sufficient emphasis and scrutiny in the literatures on the gender gap and treatment effects, with a few notable exceptions (Heckman et al., 1997; Heckman and Smith, 1998; Dehejia, 2005). Our discussions are intended to provide this emphasis. This also helps to connect the inequality literature, and the literature on gender gaps and the literature on treatment effects.

The second part and contribution of our paper is to address selection and missing wages for those who do not work. Regardless of measures, our analysis of the gender gap can be impacted by selection. 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 or actual progress. 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, mostly on the gap at the mean or median (e.g., Blau and Kahn, 2006; Olivetti and Petrongolo, 2008; Mulligan and Rubinstein, 2008).

We address the selection in our analysis at the *entire* distribution beyond the mean and median. We adopt a new quantile-Copula approach to model the joint determination of wages and participation decision for both men and women developed in Arellano and Bonhomme (2017). This approach allows us to recover the gap between the distributions of wage *offers* for the entire male and female populations.

Our account of selection also leads to a further contribution of this paper: considering value of time in measuring the gender gap. Comparing distributions of wage *offers* is informative, but for those who do not work, wage offers do not reveal "value of time" or the well-being they actually enjoy. Some individuals derive value from not working, and this is captured by their reservation wages. 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. Based on this, we provide an additional concept of the gender gap which replaces the market wages for those non-employed with their *reservation wages* instead.

[2] As is demonstrated by Theil's measures of inequality, entropies measure divergence between any income distribution and the equally distributed (or population) distribution.

This provides new results that have not been previously discussed and explored.

Using the Current Population Survey data from 1976 to 2013, we reach four main conclusions that indeed challenge the conventional perception of the gender gap. First, our baseline results (no correction for selection) provide comparability with the extant literature. Using the mean or quantile wage gaps, we confirm some of the previous findings: while women generally perform worse than men in the labor market, they are catching up with men (Blau and Kahn, 1997, Blau and Kahn, 2006, and Goldin, 2014). The gender gap has decreased over time, especially in the 1980s and early 1990s, although at a much slower rate since the mid-1990s. However, the perception of the actual gap and its dynamics varies with the measures. The quantile gaps evolve differently over the past several decades. The entropic measures provide a more nuanced picture of the evolution of the gap between wage distributions. Specifically, Generalized Entropy measures indicate a generally larger convergence until early 1990s, and a more pronounced flattening since then, for full-time workers. Moreover, the gap increases monotonically with the level of inequality aversion for entropy measures.

Second, we find selection indeed impacts all the measures of the gap and its evolution. Once selection is accounted for, convergence is slower, with a recent reversal in the trend in parts of the wage distribution between mid-1990s and the most recent recession, followed by a further marked decline in the gap, especially among low-skilled workers. The last phase is likely due to a relative deterioration in the wages of low-skilled males, rather than relative rise in wages of low skilled women. This three-phase trend is masked if selection is unaccounted for. Further, we find that (weak) uniform ranking of wage distributions between men and women is less likely, and we do not find a"uniform" narrowing of the gap at all quantiles.

Third, labor force participation varies by education and race, and this has additional impact on the gender gap for each subgroup. For example, we find that the relative economic position of less educated women lacked progress, or even deteriorated, in 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 gender gap within minority groups (blacks and Hispanics) is generally smaller than amongst whites.

Fourth, taking into account value of time, we generally find a lack of convergence. Women's relative well-being, especially among those in the upper tail, may have even worsened over time.

Our paper also has many further implications for the related literature. In addition to the main results above, our approach has allowed an examination of several other important assumptions, hypotheses and findings in the related literature. First, with estimated selection parameters, we are able to trace the evolution of selection mechanism into the 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 the difference in selection between men and women is notable, and 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 "within" inequality among both men and women has been generally increasing over time, whether or not selection is accounted for, we challenge the conventional wisdom that the increased overall (within) inequality for 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 decision theoretic bases of definitions of the gap. Section 3 discusses the data and presents the baseline results. Section 4 provides a quantile selection model examines selection results to be compared with baseline findings. In section 5., we assess the gap by education and race. In Section 6, we propose a new gender gap measure accounting for non-market (time) value for those who do not work full time. 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 or medians. More information is revealed when select percentiles are also reported.[3] Analysis of individual quantile gaps requires (unconditional) rank invariance for identification. As mentioned earlier, this requires that the male and female ranks refer to the same set of skills or at least the same intrinsic value of skills, in other words, the τ^{th} quantiles of men and women wage groups are invariant to skills, substitutions of skills, and the role of all other observed and unobserved characteristics. This issue has been discussed in e.g., Heckman et al. (1997)); tests have been proposed by Bitler et al. (2008) and Frandsen and Lefgren (forthcoming). We use the latter test with race as the additional shift variable, and reject these assumptions in all the scenarios examined in this paper. This confirms the reservations expressed by Heckman et al (1997) and others, on the ability to meaningfully identify the gender gap from individual quantile gaps.

We advocate computing an evaluation function of each distribution first, followed by the distance between the evaluated functions. This is the dominant approach in the income inequality literature that income distributions are compared and assessed over time and between groups. The average gives equal weight to all percentiles; median or any single quantile implies zero weight to other quantiles. Reporting select quantile gaps together (without explicitly specifying weights) may also invite implicit (equal) and informal weighting schemes just as averages do, and the implicit assumption of infinite substitutablity between different wage levels. There would be no explicit defense of why a particular measure (or weighting scheme) is preferred to another. Underlying each measure and evaluation function (or weighting scheme) is the decision-theoretic framework. Dalton (1920) is credited with the earliest statement of a formal correspondence between evaluation functions of distributions, like inequality measures, and "social welfare

[3] For example, in Blau and Kahn (forthcoming), the most recent, comprehensive survey of the gender gap, three select percentiles (10^{th}, 50^{th}, and 90^{th} percentiles) are examined.

functions". This framework allows us to explicitly incorporate alternative properties (such as aversion to inequality, as opposed to equal weighting) in our evaluation functions and the measurement of the gender gap.

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. A general definition of the gap is the difference of respective Evaluation Functions (EFs):

$$\text{Gap } = EF_{\gamma,\epsilon}(y^m) - EF_{\gamma,\epsilon}(y^f) \tag{1}$$

The gap at a τ^{th} quantile is $y^m_\tau - y^f_\tau$, where the median corresponds to $\tau = \frac{1}{2}$. Measures of the gap may be functions of the quantile gaps. The 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, or the mean, is a (linear) weighted function of quantile gaps. Linear functions of quantiles imply infinite substitutability of a dollar at all wage levels. Alternative functions would reflect different types of weights and/or interpersonal evaluations, reflecting degrees of aversion to inequality/dispersion. There are parallel literatures on ideal inequality (and risk) measures, and ideal entropies. The latter is summarized in Maasoumi (1993) and motivates the inequality literature.

To highlight this decision-theoretic framework, we consider an impartial observer who evaluates the wage distributions for both men and women. This person has the following Evaluation Function (EF):

$$EF_{\gamma,\epsilon} = \int_0^1 R(\tau,\gamma)U_\epsilon(y_\tau)d\tau \tag{2}$$

where $R(\tau,\gamma) = \gamma(1-\tau)^{\gamma-1}$, and $U(\cdot)$ is a concave function of wages. γ is an aversion to inequality/dispersion parameter. This class of functionals is general and underlies many conventional measures, as well as the Atkinson and S-Gini families of inequality measures (which satisfy desirable properties such as the Pigou-Dalton transfer and permutation invariance properties).[4] It allows for flexible weights at different percentiles. Holding $\gamma \neq 1$ fixed, the weight function, $R(\cdot)$, is decreasing with respect to τ, thereby assigning greater weights to lower wages in the evaluation of a wage distribution and hence measurement of the gender gap.

If only relative (scale/mean-independent) measures are to be considered, the function $U(\cdot)$ must be of the following (homothetic) form (see Pratt (1964) or Atkinson (1970)):

$$U_\epsilon(y_\tau) = \begin{cases} \frac{y_\tau^{1-\epsilon}}{1-\epsilon} & \text{if } \epsilon \neq 1 \\ \log(y_\tau) & \text{if } \epsilon = 1 \end{cases} \tag{3}$$

Note that the wage quantile y_τ itself is a special case of possible utility functions $U(\cdot)$ at $\epsilon = 0$. This leads to a linear summary function of the quantile or quantile gaps: $\int R(\tau,\gamma)(y^m_\tau - y^f_\tau)d\tau$. In the special case

[4]Invariance to permutation of individuals produces anonymity of measures with respect to identity of those who occupy a given quantile. The Pigou-Dalton transfer property (or aversion to inequality) emphasizes that one dollar reduction of gap at lower wages is relatively more valuable than one dollar at higher wages. This principle implies that any redistribution from the rich to the poor can reduce inequality. The definition of inequality-loving would be the opposite of this definition.

when $\epsilon = 0$ and $\gamma = 1$, EF is $\int y_\tau d\tau = \mathbb{E}[y]$ and the gap is the mean gap. In this case, $\gamma = 1$ (and the mean gap) implies no aversion to inequality (neutrality) in evaluating the gender gap.

A concave and increasing Evaluation Function of an impartial observer (represented by Equation 2) is known to be similarly represented as an important money metric Evaluation Function, called the Equally Distributed Equivalent (EDF) wage, given by

$$
\begin{aligned}
EDF_{\gamma,\epsilon} &= U^{-1}(EF_{\gamma,\epsilon}) & (4)\\
&= \mu_y \left(1 - I_{\gamma,\epsilon}(y)\right) & (5)
\end{aligned}
$$

where μ_y is the mean and $I_{\gamma,\epsilon}(\cdot)$ is any relative inequality measure. One can also consider alternative EFs such as those in Aaberge et al. (2013). Note that dividing both sides by the mean, we can make scale-invariant evaluations of the wage distribution based on "relative" inequality measures.

There are many inequality measures, including a monotonic transformation of the Atkinson family of inequality indices known as the Generalized Entropy (GE) family. While there exists no unique (or ideal) inequality measure, influential works by Shorrocks (1980), and Bourguignon (1979) have established the "ideal" properties of GE. These are the famed welfare properties/axioms (anonymity or invariance to permutations, continuity, scale invariance, and Pigou-Dalton principle of transfers or aversion to inequality). The inequality literature mirrors a literature on entropy functions. The shared motivation is the ability of entropy to characterize and quantify the divergence of any distribution from the uniform distribution (the case of equality having maximum entropy). For example, Theil's inequality measure is the (asymmetric Kullback-Leibler) divergence between entropy of (size) distribution of income and the population shares. Variance or log variance is also an entropy, when the underlying distribution is Gaussian. Contrasting two inequality measures mirrors the contrast between their entropies.

The contrast between the entropies of two distributions is simply a divergence measure between the densities of wages for females and males (since the uniform distributions cancel out in comparisons). Because there are competing normalizations for entropy functions, we write the symmetric GE measure of divergence between the densities of wages for females and males with a single parameter k of inequality aversion.[5]

$$
\frac{1}{2} \cdot \Big[I_k(f_1, f_2) + I_k(f_2, f_1) \Big] \quad \forall k \in [0, 1] \tag{6}
$$

where $I_k(\cdot,\cdot)$ is a GE measure of divergence given by

$$
\begin{aligned}
I_k(f_1, f_2) &= \frac{1}{k-1} \Big[\int \left(\frac{f_1}{f_2} \right)^{k-1} f_1 dy - 1 \Big] \\
I_k(f_2, f_1) &= \frac{1}{k-1} \Big[\int \left(\frac{f_2}{f_1} \right)^{k-1} f_2 dy - 1 \Big]
\end{aligned}
$$

Two popular members are noteworthy as follows:

1. The normalized and symmetrized Kullback-Leibler-Theil measure:

[5] This requires an inverse probability transformation in (2) and elsewhere.

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

2. And, at $k=\frac{1}{2}$, one obtains an entropy *distance* metric that is a normalization of the Bhattacharya-Matusita-Hellinger measure, given by:[6,7]

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

Some aspects of these entropy measures are worth noting. First, varying k is corresponding to different levels of inequality aversion in measuring the gender gap. Noting that Shannon's entropy is the basis of both the KL measure and Theil's inequality measures, the KL measure is more "inequality averse" than the Gini and the Hellinger (or S_ρ).[8] Second, these measures of the gender gap are scale invariant, and they are summary functionals of the CDF and *not affine functions* of quantiles. Finally, these measures are defined on the distribution space and are useful over many dimensions. As such, the entropy gap measures can readily accommodate multidimensional measures of the gender gap (further incorporating dimensions other than wages such as health).

2.2. Uniform Ordering: Stochastic Dominance Tests

Measures of the gender gap (discussed above) provide "complete" (cardinal) rankings. When distributions cross (especially at lower tails)[9], different measures will differ subjectively in their rankings, depending on the underlying evaluation functions. 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, all measures of the gap need to be examined relative to the underlying evaluation functions.

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:

---

[6]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 17).

[7]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.

[8]See Atkinson (1970). Gini is quite insensitive to the tails. Evaluation of policies aimed at the tails (e.g., anti poverty) often look in vain for significant movement in the Gini.

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

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

1. $Eu(y^m) \geq Eu(y^f))$ for all $u \in U_1$ with strict inequality for some u.
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. $Eu(y^m) \geq Eu(y^f)$ for all $u \in U_2$ with strict inequality for some u.
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, $\int_0^\tau y^m_u du \geq \int_0^\tau y^f_u du$ for all points on the support.

y^m FSD y^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.[10]

The SSD tests are also closely related to our decision-theoretic framework above. To see this, note that money-metric evaluations can be derived from Equation (1), and other monotonic transformations. A representation of $EF_{\gamma,\epsilon}$ (using integration by parts) reveals a very useful relation to SSD:

$$EF_{\gamma,\epsilon} = \int_0^1 \gamma(\gamma-1)(1-\tau)^{\gamma-2} GL_U(\tau)d\tau \tag{11}$$

where $GL_U(\tau) = \int_0^\tau U(y_u)du$ is the Generalized Lorenz (GL) function of $U(\cdot)$. When $U(\cdot) = y_\tau$, ranking by GL ($GL^m_U - GL^f_U = \int_0^\tau y^m_u du - \int_0^\tau y^f_u du$) is exactly the test of SSD. This has a bearing on the interpretation of the SD tests that are reported in this paper.

We want to stress that reporting the wage gap at many percentiles can be informative. We emphasize the need to be transparent about the otherwise informal subjectivity, and arbiterariness of the picture that may emerge. Even if rank invariance holds, when distributions cross, the gap changes sign, and overall statements, or impressions, of the gender gap become even more sensitive to preferences. A gap of, say

[10] 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:

$$d = \sqrt{\frac{N_1 N_2}{N_1+N_2}} \min \sup [F_m(y) - F_f(y)] \tag{9}$$

$$s = \sqrt{\frac{N_1 N_2}{N_1+N_2}} \min \sup \int_{-\infty}^{y} [F_m(t) - F_f(t)]dt \tag{10}$$

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, including older tests of quantile (inverse distribution) rankings. The latter are regarded as more difficult to implement statistically, and with comparable power.

$200, at the 10^{th} and 90^{th} percentiles are qualitatively different. We report distributions (specially after selection is counted for) that cross at low wages. This makes any meaningful SD (uniform) ranking, at any reasonable order, near impossible. Such situations require explicitly defended EF choices or weighting schemes to summarize. This is not so when we find FSD or SSD. Our proposed entropy measures do not "solve" the conundrum so eloquently put forth by Arrow's impossibility theorems. Our discussions instead expose this challenging situation by showing what can and cannot be done, even with the metric member of a very flexible EF family.

3. Baseline Results

3.1. Data

We examine the period 1976-2013, March Current Population Survey (CPS) (Flood et al., 2017). 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.[11]

Our sample includes individuals aged between 18 and 64 who 1) work only for wages and salary, 2) do not live in group quarters, 3) work more than 20 weeks (inclusive), and more than 35 hours per week in the previous year (e.g., Mulligan and Rubinstein 2008). Our baseline results focus on unconditional distributions ignoring selection. Information about sample size is provided in the supplemental material (Table E.1). [12]

To achieve some economy in prose, throughout this paper, we refer to this full-time working sample as W, and to the Selection Corrected results as SC.

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

This section is intended to facilitate comparison with the existing literature. Tables (1) and (2) report a number of measures of the gender gap. Columns (1) and (6) of Table (1) 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. Table (2) displays entropy measures of the gender gap with varying levels of inequality aversion, $k = 0.1, \ldots, 0.9$. S_ρ is further normalized, taking values in $[0, 1]$, and all entropy measures are multiplied by 100 to facilitate the presentation. Note that $S_\rho = 2 \times I_k$ when $k = 1/2$.

Our baseline results confirm three important findings in the literature. First, the gap is substantial and positive, but heterogenous across the distribution. In 1976, for instance, the gender gap at the 10^{th} percentile is about 31 percentage points, about 50 percentage points at the 90^{th} percentile, and varies

[11] 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).

[12] We use person-level weight (WTSUPP) variable throughout our analysis (which "should be used in analyses of individual-level CPS supplement data"). See IPUMS-CPS website for more details (Flood et al., 2017). We also repeat all our analysis using NBER Outgoing Rotation Group data to assess the robustness of our results (See Appendix B), as suggested by a referee.

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

Year	Mean (1)	10th (2)	25th (3)	50th (4)	75th (5)	90th (6)
1976	0.432	0.311	0.427	0.461	0.461	0.486
1977	0.419	0.297	0.397	0.470	0.465	0.470
1978	0.425	0.272	0.392	0.465	0.472	0.465
1979	0.421	0.262	0.386	0.468	0.487	0.474
1980	0.408	0.252	0.382	0.446	0.475	0.448
1981	0.395	0.263	0.338	0.447	0.477	0.440
1982	0.394	0.254	0.345	0.448	0.467	0.443
1983	0.376	0.240	0.323	0.427	0.439	0.451
1984	0.356	0.247	0.298	0.412	0.427	0.405
1985	0.343	0.208	0.288	0.357	0.401	0.414
1986	0.327	0.189	0.297	0.366	0.387	0.393
1987	0.322	0.214	0.288	0.376	0.365	0.378
1988	0.305	0.222	0.264	0.329	0.353	0.354
1989	0.298	0.210	0.244	0.323	0.324	0.352
1990	0.283	0.183	0.265	0.303	0.336	0.324
1991	0.254	0.143	0.208	0.273	0.307	0.330
1992	0.241	0.107	0.211	0.251	0.285	0.299
1993	0.228	0.145	0.194	0.262	0.274	0.285
1994	0.215	0.142	0.175	0.235	0.262	0.281
1995	0.221	0.143	0.197	0.239	0.283	0.276
1996	0.231	0.148	0.186	0.240	0.248	0.264
1997	0.227	0.158	0.210	0.241	0.262	0.259
1998	0.228	0.131	0.214	0.261	0.250	0.274
1999	0.235	0.154	0.192	0.231	0.251	0.262
2000	0.232	0.182	0.202	0.247	0.285	0.284
2001	0.226	0.143	0.192	0.227	0.261	0.300
2002	0.216	0.117	0.185	0.205	0.236	0.278
2003	0.208	0.125	0.147	0.203	0.238	0.265
2004	0.188	0.130	0.153	0.164	0.230	0.288
2005	0.190	0.105	0.182	0.180	0.219	0.255
2006	0.191	0.113	0.145	0.182	0.204	0.243
2007	0.180	0.125	0.153	0.183	0.206	0.227
2008	0.172	0.118	0.103	0.192	0.214	0.229
2009	0.186	0.136	0.148	0.163	0.212	0.270
2010	0.187	0.131	0.134	0.173	0.220	0.254
2011	0.175	0.111	0.140	0.173	0.211	0.236
2012	0.178	0.111	0.139	0.170	0.215	0.266
2013	0.167	0.108	0.113	0.172	0.198	0.223

Table 2: Entropy Measures of The Gender Gap (fulltime employed)

Year	S_ρ (1)	Theil (2)	$k=0.1$ (3)	$k=0.2$ (4)	$k=0.3$ (5)	$k=0.4$ (6)	$k=0.5$ (7)	$k=0.6$ (8)	$k=0.7$ (9)	$k=0.8$ (10)	$k=0.9$ (11)
1976	10.566	44.460	4.384	8.618	12.788	16.945	21.137	25.417	29.840	34.474	39.452
1977	10.110	42.654	4.173	8.231	12.223	16.200	20.210	24.301	28.521	32.924	37.561
1978	10.247	43.238	4.240	8.354	12.397	16.424	20.486	24.636	28.927	33.415	38.160
1979	10.127	42.346	4.184	8.249	12.250	16.235	20.254	24.353	28.583	32.996	37.660
1980	9.857	40.874	4.071	8.026	11.926	15.814	19.731	23.721	27.828	32.105	36.637
1981	9.221	38.846	3.807	7.505	11.150	14.783	18.444	22.175	26.018	30.022	34.260
1982	8.823	37.094	3.653	7.196	10.679	14.147	17.646	21.220	24.917	28.786	32.877
1983	7.645	31.984	3.144	6.212	9.238	12.253	15.290	18.380	21.554	24.848	28.297
1984	6.747	28.230	2.782	5.485	8.156	10.821	13.504	16.231	19.031	21.940	25.042
1985	6.258	25.952	2.574	5.083	7.561	10.031	12.519	15.047	17.642	20.334	23.168
1986	5.607	23.100	2.291	4.539	6.764	8.985	11.217	13.477	15.784	18.155	20.619
1987	5.023	20.584	2.043	4.056	6.052	8.044	10.044	12.066	14.122	16.225	18.387
1988	4.526	18.542	1.844	3.659	5.459	7.254	9.058	10.882	12.737	14.636	16.593
1989	4.215	17.458	1.717	3.406	5.082	6.754	8.434	10.132	11.858	13.625	15.454
1990	3.645	14.836	1.483	2.941	4.389	5.835	7.287	8.753	10.241	11.764	13.345
1991	3.109	12.710	1.260	2.506	3.743	4.979	6.219	7.468	8.735	10.024	11.342
1992	2.784	11.280	1.125	2.239	3.347	4.454	5.564	6.681	7.810	8.955	10.126
1993	2.518	10.232	1.020	2.029	3.033	4.035	5.041	6.053	7.077	8.117	9.182
1994	2.153	8.734	0.864	1.726	2.584	3.440	4.298	5.160	6.028	6.903	7.776
1995	2.048	8.314	0.825	1.644	2.461	3.276	4.093	4.914	5.741	6.577	7.424
1996	2.165	8.942	0.888	1.752	2.613	3.472	4.336	5.208	6.096	7.009	7.988
1997	2.071	8.418	0.852	1.681	2.507	3.334	4.163	5.001	5.851	6.726	7.670
1998	2.082	8.408	0.844	1.679	2.510	3.340	4.173	5.011	5.856	6.715	7.599
1999	2.236	9.232	0.910	1.807	2.695	3.582	4.473	5.373	6.289	7.226	8.193
2000	2.059	8.350	0.836	1.660	2.480	3.300	4.122	4.950	5.787	6.638	7.522
2001	1.913	7.832	0.769	1.536	2.299	3.059	3.822	4.589	5.363	6.145	6.925
2002	1.781	7.352	0.714	1.425	2.132	2.839	3.546	4.258	4.976	5.700	6.423
2003	1.684	6.868	0.681	1.354	2.022	2.689	3.358	4.033	4.718	5.417	6.133
2004	1.353	5.374	0.523	1.061	1.596	2.130	2.663	3.195	3.724	4.243	4.709
2005	1.393	5.772	0.562	1.120	1.675	2.229	2.784	3.343	3.907	4.479	5.055
2006	1.321	5.262	0.538	1.067	1.592	2.116	2.642	3.173	3.714	4.268	4.843
2007	1.102	4.358	0.446	0.887	1.327	1.768	2.208	2.651	3.097	3.549	4.015
2008	1.149	4.568	0.477	0.938	1.398	1.857	2.319	2.786	3.261	3.754	4.297
2009	1.275	5.136	0.515	1.025	1.534	2.043	2.552	3.064	3.580	4.102	4.637
2010	1.275	5.206	0.522	1.033	1.542	2.052	2.563	3.078	3.599	4.131	4.696
2011	1.118	4.550	0.459	0.908	1.355	1.802	2.251	2.703	3.161	3.630	4.135
2012	1.129	4.492	0.457	0.907	1.357	1.807	2.258	2.711	3.167	3.630	4.109
2013	0.974	3.988	0.414	0.807	1.199	1.593	1.989	2.389	2.799	3.228	3.727

[1] Note that all entropy measures are multiplied by 100. The original values S_{rho} are normalized to be between 0 and 1. k correspond to varying levels of inequality aversions.

between 43 and 46 percentage points at other percentiles. Second, 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. However, 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.[13] Finally, the impression of the cyclicality of the overall gap varies across measures. For example, In 2009 (the great recession), the average gender gap increased, while the median gap decreased! The 10^{th} percentile gap fluctuates more than at other quantiles; the gender gap at the 90^{th} percentile exhibits a consistently declining trend with some fluctuation.

The entropy measures corroborate some of the prior findings. Both S_ρ and Theil (entropy) measures are statistically significantly different from zero in all cases, indicating sizable gender gap. However, these measures also provide a different impression of how the gender gap fluctuates over time, relative to the conventional measures. Accounting for *not only* increasing earnings *but also* the greater dispersion (inequality increasing) accompanying it, the S_ρ measure decreased in 1977, consistent with the decrease at all quantiles except at the median and 70^{th} percentile; it increased in 1999, consistent with only the 10th and 75th percentiles. In the most recent recession (2009), when the change in the mean gap differs in direction from the median gap, the S_ρ agrees with the mean. The KL and other entropy measures of the gap generally agree with the metric entropy measure. But we can see that the gender gap increases monotonically with the level of inequality aversion.

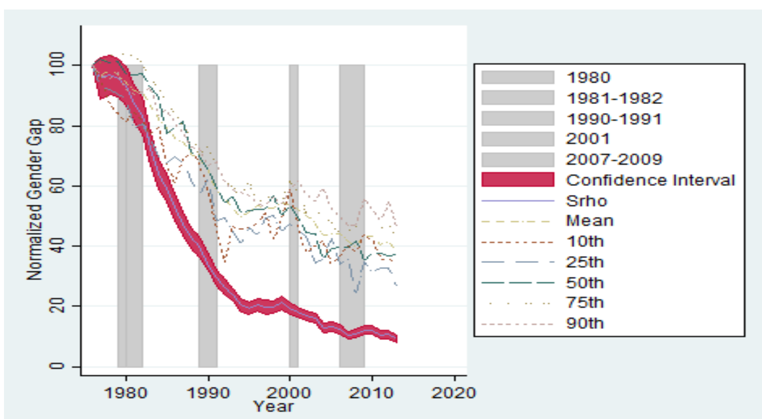


Figure 1: The Trend of Gender Gap (Shaded Areas correspond to the recession periods announced by NBER)

To further contrast how the gender gap measures respond to the business cycles, we plot the normalized gap measures in Figure (1), with confidence intervals for S_ρ in (red) shaded areas.[14] Conventional measures of the gap correspond differently to business cycles, with varying directions. The entropy gap is relatively robust to recessions, but more responsive to the recession in the most recent recession. This may reflect the fact that in such worsened economic conditions, greater changes in the gap in lower tail occurred,

[13]All these results are also conveniently depicted by patterns of changes in different measures in Table (E.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.

[14]We normalize the measures in Tables (1) and (2) 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).

and such changes are reflected in the entropy measures that account for inequality and integrates different movements of the gap at different quantiles.

3.2.1. The Long-Run Trend Implied by Various Measures

To better visualize and contrast the long-run trends implied by these measures, in Figure (2) and (3) we present smoothed trend lines for each measure using lowess (i.e., a locally weighted regression of measures on time). The movements implied by conventional measures have been noted in 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 (forthcoming).

Table 3: 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	-0.030	-0.026	-0.024
1994-2013	-0.051	-0.051	-0.019	-0.016	-0.028	-0.023	-0.017	-0.012

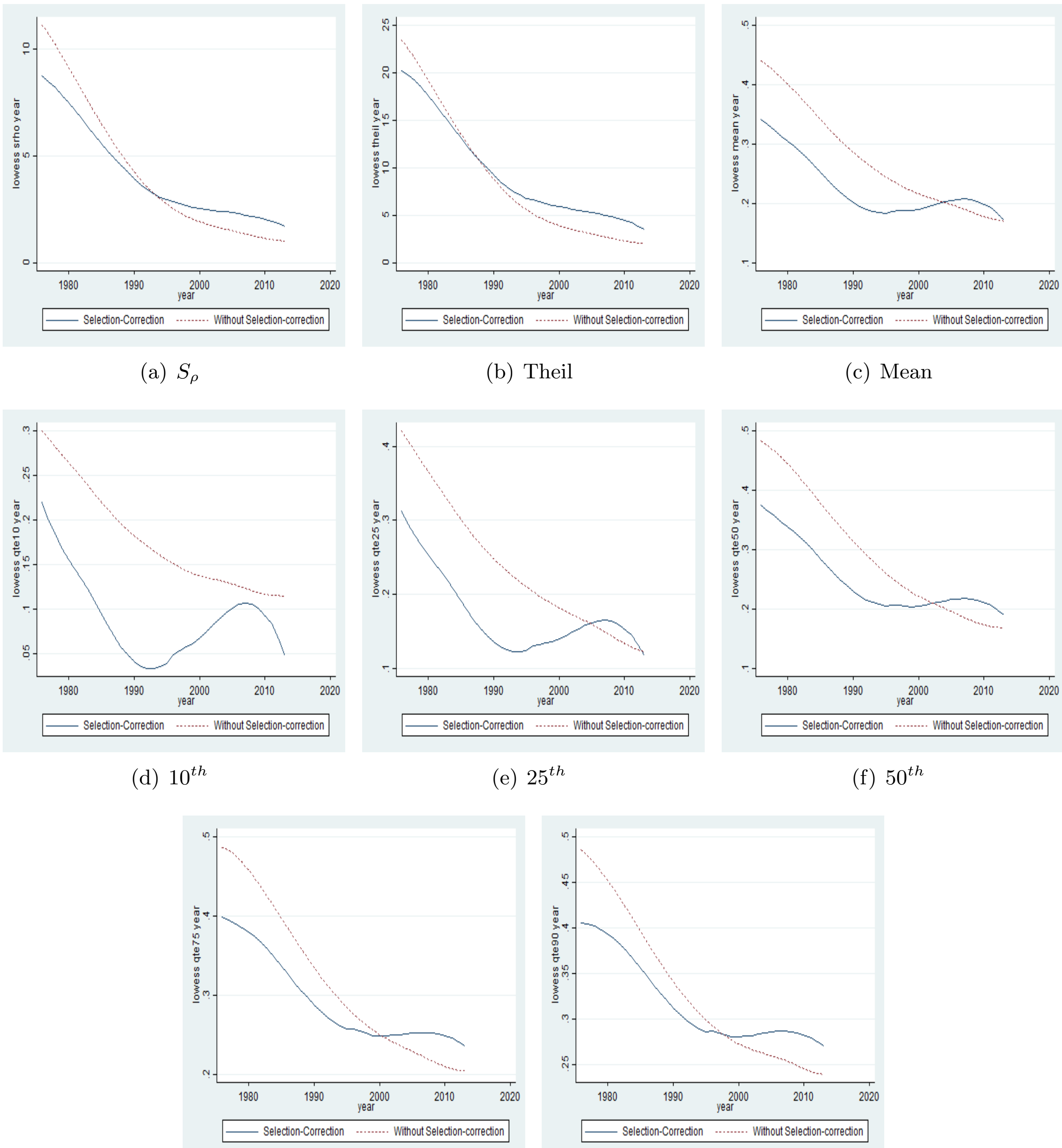
[1] These values are long-run compound annual change rates implied by the initial and the last smoothed values of each period.

The long-run trend implied by our entropy measures are nearly the same and generally agree with conventional measures. Since the units for these measures are not directly comparable, Table (3) quantifies the differences in the implied long-run trends, reporting implied compound annual convergence rates over the entire period. Additional entropy measures with varying degrees of inequality aversions show nearly identical patterns. The results can be found in Table (A.3) of the supplemental material.[15,16,17] The entropy measures suggest much larger declines and rates of decline: the gender gap dropped *precipitously* before 1990s, but the trend of "convergence" slowed down since 1990s. The entropy measures decline by

[15]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.

[16]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.

[17]**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 (E.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 at the mean. For the data here the discrepancies are relatively modest. However, it is conceivable that in other contexts such differences may be large.



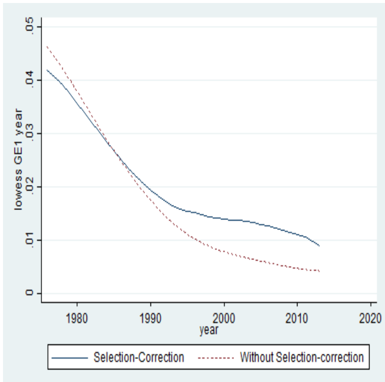
(a) S_ρ (b) Theil (c) Mean

(d) 10^{th} (e) 25^{th} (f) 50^{th}

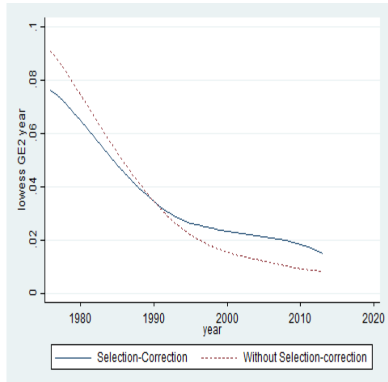
(g) 75^{th} (h) 90^{th}

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

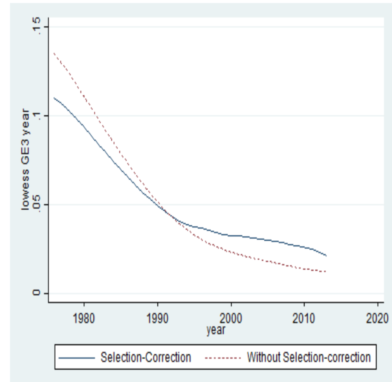
Note: The selection correction method and the corresponding results, in which we consider men and women who do not work full-time, will be discussed in detail below. In the interest of space, however, we present these comparisons here.



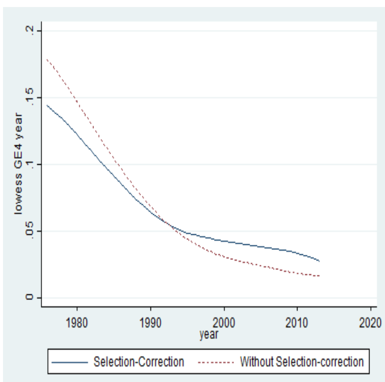
(a) $k = 0.1$



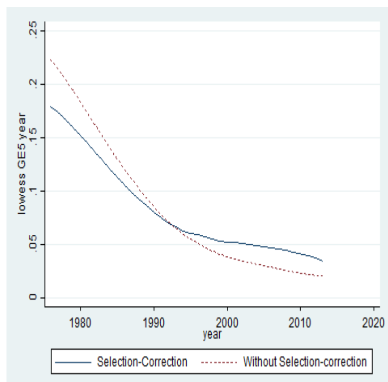
(b) $k = 0.2$



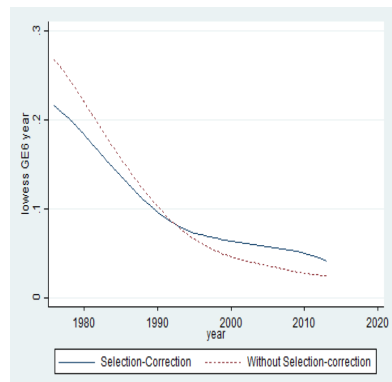
(c) $k = 0.3$



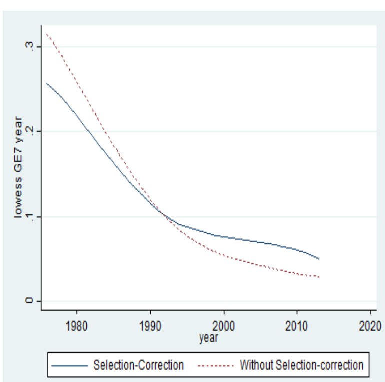
(d) $k = 0.4$



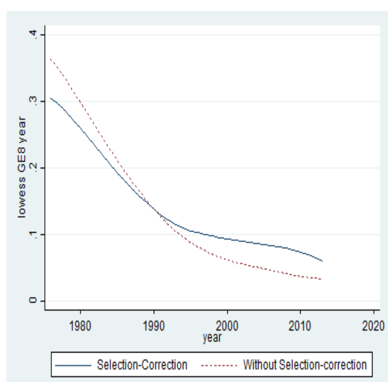
(e) $k = 0.5$



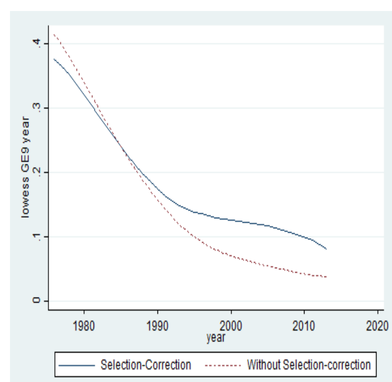
(f) $k = 0.6$



(g) $k = 0.7$



(h) $k = 0.8$



(i) $k = 0.9$

Figure 3: Comparison of Smoothed Trend of The Gender Gap with and without Selection Correction for Entropy Measures with Varying Degree of Inequality Aversions (excluding 2010)

Note: The selection correction method and the corresponding results, in which we consider men and women who do not work full-time, will be discussed in detail below. In the interest of space, however, we present these comparisons here.

about 90 percent, other measures by about 50 percent, over the entire period.

3.2.2. Stochastic Dominance Rankings

We examine the unconditional empirical distributions of earnings for men and women. Baseline SD rankings are similar across years. We exemplify these results with the most recent year in Table (4). 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.[18]

Table 4: 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 Stochastic Doninance (FSD) in every year. Ranking in 2013 is representative.

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 SSD and 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*. We also repeat all of our baseline analysis using NBER Outgoing Rotation Group data and find strikingly similar results (see Appendix B in the supplemental material).

However, these baseline results and the findings from the existing literature are not sustained once selection is accounted for. They also change when we report counterfactual results with controls for some characteristics. We now turn to the second part of our paper that examines the impact of selection on the gender gap.

4. Selection Corrected (SC) Analysis

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 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, the relatively higher 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%.

[18] 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.

If wages for non-working women are systematically lower than for full-time workers (positive selection), the baseline (W) results would under-estimate the gender gap, while stochastic dominance rankings can remain unchanged and even strengthened. If the wages of non-working women are higher than full-time workers (negative selection), the baseline results above would over-state 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 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 has been adjusted for selection. Recovery of marginal distribution from conditional quantiles is addressed in Machado and Mata (2005). The mathematical rationale for the two-step method is briefly summarized in Appendix D in the supplemental material. The asymptotic statistical theory recently developed in Chernozhukov et al. (2013) may be applicable. Our procedure differs from Machado and Mata (2005) and some subsequent analysis in that we take into account selection when estimating the conditional quantiles.[19] Once the marginal distribution is obtained, the calculations of the gender gap measures, as well as well as stochastic dominance tests are straightforward to implement

There are a few methods in the literature to *point* identify conditional quantiles with selection. An approach proposed in Arellano and Bonhomme (2017) 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

[19] 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 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, or the pdf with robust nonparametric kernel density estimation on the data.

distribution and the participation decision, leaving a good deal of flexibility for marginal processes for wage and selection. We do not impose the restriction in Blundell et al. (2007), while maintaining some commonly invoked assumptions as detailed below. 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 above.

4.1.1. 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 including unobserved wages for non workers.

To this end, we adopt a quantile-copula function approach proposed in Arellano and Bonhomme (2017). A more detailed survey of possible alternative methods and detailed reasons for our choice are moved to footnotes (20) and (21). This approach, by adding (slightly) more structure and hence information, can address selection and enables identification. It also has greater flexibility in modelling the joint dependence of the marginal variables. In the presence of selection, their approach entails shifting the percentiles as a function of the amount of selection.[20,21]

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

$$\ln(w) = g(x, u) \quad u|x \sim Uniform(0, 1) \tag{12}$$

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] Unobservables, u, are the rank variable or quantile and thus can be fixed in estimations. The participation decision written in a normalized form is

[20] 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 (2017) 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.

[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 (2017) and Arellano 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).

[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 (12). The second follows from the fact that conditional on x, u is uniformly distributed.

given by:

$$S = I(v \leq p(z)) \quad v|x \sim Uniform(0,1) \tag{13}$$

where $p(z) = \Pr[S = 1|z]$ is the propensity score, and assuming $p(z) > 0$ with probability one.[23] $I(\cdot)$ is an indicator function (equal to one if the argument is true, zero otherwise). Let $z = (x', \tilde{z}')'$, where $\tilde{z}$ includes a vector of IVs statistically independent of both (u, v) given x. An exclusion restriction is through a variable that affects the selection equation only (see below).

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 idea proposed by Arellano and Bonhomme (2017). [24] Given (a) availability of an exclusion restriction, (b) absolutely continuous bivariate distribution of (U, V) (represented by its copula, $C(u, v)$), (c) continuous outcome, and (d) $p(z) > 0$, $g(\cdot)$ is nonparametrically identified.

4.1.2. Practical Implementation

Following the literature, we work with a linear conditional quantile function $g(x, u) = x'\beta(u)$. This is *nonlinear* since it allows x to have differential impact at different quantiles.[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. Linear quantile regression can provide a weighted least squares approximation to an unknown and potentially nonlinear conditional quantile regression, Angrist et al. (2006). Below we provide some graphic evidence of the performance of such linear non-separable models. 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 (forthcoming) refer to as "human capital specification". [26]

[23]Note that Equation (13) is a normalization commonly used in the treatment effects literature. Note that $\mathbb{E}[S|z] = \Pr[S = 1|z] = p(z) = \mathbb{E}[S = 1|p(x)] = \Pr[S = 1|p(z)]$.

[24]The algorithm is provided in detail in Appendix C 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]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.

Propensity scores are estimated by probit models with a flexible specification that includes polynomial terms of the continuous variables up to third order, as well as interaction terms between them and other discrete variables, in addition to the IV. A linear index model is employed here, and these variables enter the probit model additively. The main exclusion restriction is the presence/number of young children (the reasons and justifications will be discussed below). Note that the set of variables do not completely overlap with those in the wage equation, thereby providing some additional "exclusion restrictions" for identification.

Identification is further aided by the copula function. While identification analysis in Arellano and Bonhomme (2017) is general and covers the case where the copula is nonparametric, we adopt their choice of Frank copula with 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, determined by the data. We will verify its performance and assess its robustness below.

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. This parameter will also be useful to assess the robustness of our results (see below).

There are three-steps of implementations, details of which are provided in the supplemental material: 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 $\tau = 0.02, 0.03, \ldots, 0.97, 0.98$.[27]

4.2. Results with Selection Correction (SC)

We first examine underlying assumptions and robustness of our approach to addressing selection. This is followed by estimates of the gender gap and SD tests.

4.2.1. Validity of Assumptions and Robustness of Results

Our approach is premised on two important methodological choices and assumptions, the choice of instrument and an exclusion restriction, and the copula function. We offer arguments in support of our choices, the validity of these assumptions, and the robustness of our results when the assumption validity cannot be directly tested. We show that our assumptions are not violated and the results are relatively robust to alternatives.

IV and the Exclusion Restriction We follow the tradition of using the presence/number of children under age 5 as an IV. 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

[27]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.

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. Theoretical discussions and justifications in favor of our choice are moved to Footnote (28).[28]

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.

The second set of results is concerned with the independence of the excluded IV and potential wages (at least conditional on X). This is the main concern in empirical analysis, and instead of analyzing alternative (controversial) exclusion restrictions, we provide a more rigorous statistical test. We use the method proposed in Huber and Mellace (2014). They show that under our model assumptions, the following inequalities hold

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

Such inequalities imply the following null hypotheses:

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

Huber and Mellace (2014) test the validity of IVs 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.

Copula-Model Fit and Specification Errors Identification is mainly achieved through exclusion restrictions, but also aided by the copula function, especially for the results among men. We assess how well the quantile selection models perform under these assumptions. We can do this since we are able to identify

[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

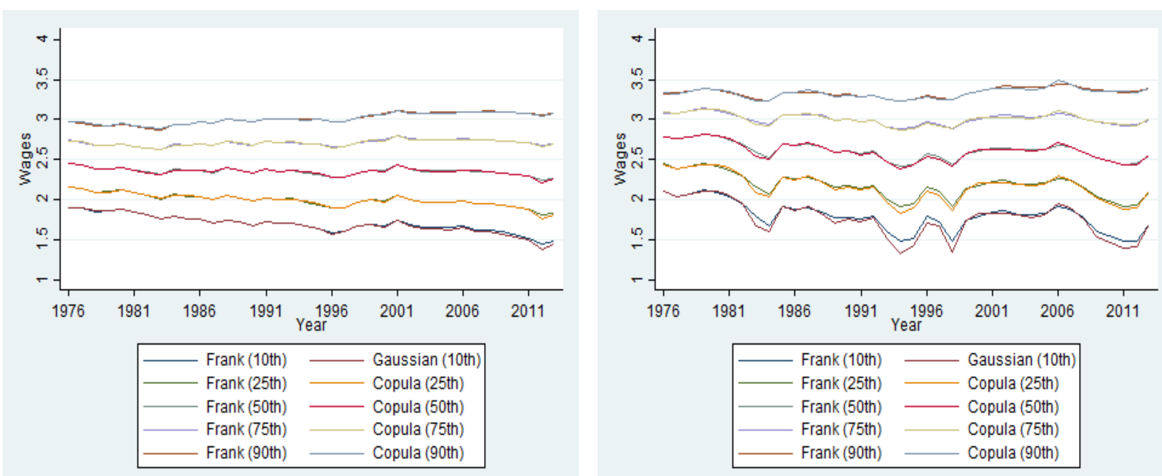
[29]Note that this test can be readily extended to the multivalued case, but for ease of exposition and computation, we consider a binary case here, i.e., the presence of young children. A negative test statistic with a large p-value indicates that IV validity is *not* violated. Readers are referred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

and compare the wage distributions of the full-time workers constructed two different ways. One based on quantile selection model, the other is the corresponding 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 only for women (see, also, footnote (31) for more discussions of these early results).[30]

Robustness to Choice of Copula Frank copula is a low-dimensional copula with appealing features as discussed above. We assess the robustness of our results by re-estimating our models using a Gaussian copula, which is another low-dimension Copula, and more importantly, provides dependence parameters that could be compared to ρ by the implied Spearman correlation coefficient. Note that in the special case when both marginal distributions of u and v are normal, the copula is a bivariate normal distribution, as in the Heckman model (Lee, 1983). Our Gaussian-copula specification is based on arbitrary marginals and hence more general.

We first compare the wage distributions for both men and women recovered by models based on different copulas in Figure (4). These estimates are nearly identical. We also calculate and compare the Spearman correlation coefficients for both copula functions in Table (5), and we again find that they are similar and nearly identical in many cases. The robustness of our results to alternative copulas may be due to the flexibility of our copula function in which the marginals can be completely nonparametric.



(a) Female

(b) Male

Figure 4: Wage Distributions based on Frank vs Gaussian Coupulas

[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.

Table 5: Spearman Correlation Coefficients for Different Copula models Its Corresponding Signs

Year	Frank Copula		Gaussian Copula		Year	Frank Copula		Gaussian Copula	
	Female	Male	Female	Male		Female	Male	Female	Male
	(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)
1976	0.34	0.12	0.34	0.12	1995	-0.10	-0.66	-0.11	-0.73
1977	0.28	-0.02	0.29	0.00	1996	-0.18	-0.12	-0.21	-0.31
1978	0.17	0.05	0.18	0.04	1997	-0.18	-0.39	-0.20	-0.44
1979	0.15	0.17	0.14	0.14	1998	-0.11	-0.73	-0.11	-0.82
1980	0.23	0.14	0.24	0.20	1999	-0.09	-0.37	-0.08	-0.40
1981	0.23	0.16	0.24	0.22	2000	-0.12	-0.25	-0.17	-0.21
1982	0.18	-0.05	0.18	-0.05	2001	-0.03	-0.24	-0.04	-0.30
1983	0.09	-0.32	0.09	-0.50	2002	-0.18	-0.19	-0.21	-0.31
1984	0.18	-0.49	0.18	-0.58	2003	-0.28	-0.29	-0.31	-0.32
1985	0.11	0.00	0.12	-0.02	2004	-0.26	-0.28	-0.29	-0.33
1986	0.10	-0.08	0.10	-0.11	2005	-0.24	-0.19	-0.27	-0.27
1987	-0.04	-0.03	-0.03	-0.03	2006	-0.18	0.02	-0.21	0.12
1988	0.06	-0.20	0.07	-0.24	2007	-0.25	-0.09	-0.30	-0.09
1989	0.02	-0.42	0.03	-0.47	2008	-0.30	-0.41	-0.31	-0.40
1990	-0.10	-0.39	-0.11	-0.41	2009	-0.31	-0.56	-0.34	-0.62
1991	0.04	-0.34	0.04	-0.41	2011	-0.39	-0.65	-0.41	-0.70
1992	0.00	-0.21	0.00	-0.22	2012	-0.51	-0.63	-0.55	-0.70
1993	-0.01	-0.55	-0.01	-0.64	2013	-0.41	-0.33	-0.45	-0.31
1994	-0.07	-0.69	-0.09	-0.79					

[1] We find First-order Stochastic Doninance (FSD) in every year. Ranking in 2013 is representative.

While the evidence is not definitive and more robustness assessment is warranted, it does provide increased confidence in the findings that we now turn to.

4.2.2. Selection and The Magnitude of the Gap

Time-varying Selection Patterns Tables (6) and (7) display the segments in the "true" wage distribution containing the non-full-time men and women, respectively. Specifically, we categorize individuals into 10 groups by quantiles in their respective true wage distribution, and examine the percentage of each category among non-full-time workers. While non-full-time are generally from lower tails of the wage distribution, this is not monotonic. The variation is better revealed when we look at 100 categories. Moreover, the share of non-full time from the lower tails has an increasing trend, while the share of those from the upper tails exhibits a decreasing trend. Over time, more higher-wage earners entered the labor market than lower-wage earners. This pattern is consistent with the changes in the dependence parameter, ρ. The latter can also help understand how selection pattern changes once we control for individual characteristics. The estimated dependence parameter, ρ, is given in Tables (8) and (9), for women and men 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). 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 in Section (7.3) and discuss the other in footnote (40). For men, on the other hand, the parameter is consistently negative throughout, except for a few early years, which suggests positive selection.

Selection-corrected Gender Gap: Magnitudes, Cyclicality, and Heterogeneity 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.[31] However, the direction of the changes is unclear, a priori, when selection is accounted for *both* men *and* women.

Results are presented in Tables (10) and (11) . Corresponding standard errors are provided in Table (E.30) in the supplemental material. Regardless of measures, the selection corrected (SC) gender gap

[31] 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 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 for the same period, ranging from .330 to .384.

Table 6: The Portion of the "True" Wage Distribution the Non-Full-Time Working People are From (Female)

	Quantile Groups									
Year	1	2	3	4	5	6	7	8	9	10
1976	10.43	10.36	10.26	10.16	10.03	9.96	9.87	9.77	9.67	9.51
1977	10.44	10.34	10.25	10.13	10.07	9.96	9.88	9.79	9.71	9.42
1978	10.49	10.41	10.27	10.15	10.04	9.97	9.84	9.75	9.60	9.49
1979	10.45	10.41	10.27	10.14	10.06	9.98	9.82	9.74	9.64	9.49
1980	10.58	10.46	10.28	10.16	10.03	9.94	9.86	9.72	9.64	9.34
1981	10.80	10.47	10.30	10.13	10.04	9.91	9.82	9.71	9.61	9.23
1982	10.78	10.54	10.32	10.21	10.06	9.95	9.81	9.63	9.44	9.27
1983	11.01	10.63	10.44	10.28	10.13	9.94	9.73	9.58	9.34	8.92
1984	11.08	10.67	10.41	10.22	10.06	9.84	9.72	9.56	9.43	9.02
1985	11.27	10.73	10.45	10.23	10.06	9.92	9.71	9.44	9.29	8.88
1986	11.39	10.70	10.44	10.24	10.09	9.92	9.68	9.45	9.22	8.88
1987	11.43	10.83	10.54	10.26	10.04	9.91	9.68	9.42	9.21	8.67
1988	11.46	10.84	10.51	10.21	10.03	9.86	9.73	9.51	9.17	8.67
1989	11.57	10.92	10.62	10.36	10.11	9.82	9.61	9.35	9.06	8.57
1990	11.64	10.95	10.59	10.38	10.16	9.90	9.66	9.34	8.97	8.40
1991	11.78	10.93	10.61	10.36	10.07	9.84	9.57	9.32	9.04	8.46
1992	11.96	11.12	10.72	10.38	10.05	9.77	9.53	9.22	8.89	8.35
1993	12.13	11.16	10.76	10.43	10.07	9.73	9.43	9.18	8.84	8.28
1994	12.14	11.04	10.70	10.38	10.12	9.83	9.50	9.16	8.88	8.25
1995	12.27	11.14	10.73	10.34	10.04	9.71	9.44	9.13	8.85	8.36
1996	12.43	11.22	10.74	10.34	10.02	9.75	9.34	9.05	8.79	8.33
1997	12.36	11.24	10.83	10.44	9.97	9.70	9.31	9.04	8.77	8.34
1998	12.15	11.16	10.81	10.44	10.05	9.73	9.41	9.11	8.80	8.35
1999	12.15	11.16	10.75	10.39	9.95	9.70	9.44	9.15	8.88	8.43
2000	12.21	11.07	10.70	10.37	10.12	9.74	9.40	9.13	8.86	8.38
2001	12.00	11.05	10.58	10.23	9.94	9.70	9.43	9.23	9.08	8.77
2002	12.28	11.19	10.66	10.30	9.95	9.69	9.40	9.14	8.87	8.52
2003	12.42	11.11	10.62	10.23	9.96	9.63	9.35	9.11	8.92	8.65
2004	12.38	11.11	10.68	10.25	9.93	9.59	9.39	9.14	8.90	8.65
2005	12.37	11.22	10.78	10.25	9.96	9.59	9.38	9.10	8.81	8.55
2006	12.54	11.20	10.71	10.25	9.91	9.64	9.34	9.06	8.84	8.52
2007	12.50	11.18	10.71	10.22	9.92	9.61	9.34	9.07	8.88	8.57
2008	12.75	11.33	10.77	10.31	9.90	9.50	9.26	8.99	8.74	8.46
2009	12.90	11.38	10.79	10.33	9.92	9.56	9.23	8.93	8.63	8.33
2010	15.47	12.39	10.04	9.99	10.60	8.98	9.03	7.10	7.28	9.13
2011	13.46	11.54	10.79	10.36	9.91	9.45	9.09	8.77	8.50	8.14
2012	13.40	11.54	10.96	10.35	10.02	9.52	9.13	8.71	8.42	7.94
2013	13.38	11.49	10.81	10.43	9.96	9.57	9.16	8.79	8.46	7.95

Table 7: The Portion of the "True" Wage Distribution the Non-Full-Time Working People are From (Male)

	Quantile Groups									
Year	1	2	3	4	5	6	7	8	9	10
1976	15.36	12.60	11.10	10.01	9.41	8.83	8.53	8.25	8.07	7.86
1977	15.30	12.38	10.93	9.91	9.32	8.94	8.56	8.40	8.20	8.08
1978	14.70	11.94	10.82	9.96	9.41	8.94	8.81	8.58	8.49	8.34
1979	14.00	11.70	10.61	9.91	9.38	9.09	8.90	8.81	8.76	8.84
1980	13.95	11.65	10.56	9.92	9.37	9.11	8.96	8.83	8.74	8.91
1981	14.12	11.77	10.73	9.93	9.39	9.06	8.86	8.74	8.71	8.69
1982	13.99	11.74	10.65	9.89	9.38	9.14	8.95	8.79	8.75	8.73
1983	14.80	11.85	10.70	9.91	9.41	9.04	8.79	8.60	8.53	8.38
1984	14.81	11.97	10.78	9.94	9.40	9.05	8.73	8.56	8.43	8.33
1985	14.17	11.57	10.59	9.91	9.39	9.05	8.88	8.83	8.76	8.83
1986	14.38	11.86	10.72	9.99	9.41	9.07	8.83	8.64	8.58	8.52
1987	14.27	11.79	10.68	9.87	9.39	9.03	8.92	8.82	8.69	8.53
1988	14.07	11.57	10.56	9.92	9.46	9.10	8.92	8.83	8.83	8.74
1989	15.02	11.65	10.51	9.85	9.33	9.09	8.82	8.68	8.53	8.52
1990	14.93	11.67	10.48	9.85	9.34	8.98	8.79	8.65	8.61	8.71
1991	14.63	11.61	10.61	9.83	9.39	9.04	8.78	8.69	8.67	8.74
1992	14.52	11.77	10.60	9.87	9.33	8.96	8.78	8.70	8.67	8.78
1993	15.57	11.69	10.63	9.89	9.35	8.95	8.66	8.48	8.37	8.42
1994	15.70	11.67	10.67	9.95	9.37	9.00	8.64	8.44	8.33	8.23
1995	16.15	11.80	10.61	9.87	9.27	8.86	8.51	8.39	8.28	8.25
1996	14.73	11.71	10.53	9.73	9.31	8.98	8.78	8.74	8.69	8.82
1997	15.55	11.73	10.55	9.79	9.26	8.85	8.64	8.50	8.45	8.67
1998	15.92	11.91	10.67	9.94	9.26	8.85	8.57	8.38	8.26	8.24
1999	14.99	11.76	10.73	9.98	9.31	9.01	8.73	8.61	8.44	8.44
2000	14.56	11.77	10.70	9.94	9.43	9.07	8.82	8.68	8.55	8.47
2001	15.49	12.22	10.81	9.93	9.29	8.87	8.60	8.46	8.19	8.13
2002	15.50	12.31	10.92	9.99	9.33	8.94	8.63	8.34	8.09	7.95
2003	16.01	12.30	10.84	9.89	9.32	8.92	8.62	8.32	8.05	7.72
2004	15.62	12.25	11.00	9.96	9.36	8.95	8.61	8.34	8.10	7.82
2005	15.35	12.18	10.91	10.01	9.45	8.97	8.67	8.42	8.18	7.87
2006	14.78	12.21	10.87	9.93	9.41	9.03	8.78	8.55	8.29	8.14
2007	14.84	12.13	10.84	10.09	9.46	9.12	8.80	8.55	8.23	7.92
2008	15.80	12.35	10.94	10.02	9.43	8.96	8.62	8.32	8.00	7.56
2009	16.43	12.41	11.11	10.16	9.47	9.00	8.47	8.11	7.73	7.12
2010	16.29	12.37	11.15	10.35	9.70	9.08	8.59	8.12	7.58	6.77
2011	16.63	12.50	11.18	10.25	9.56	8.99	8.47	8.07	7.51	6.84
2012	16.91	12.69	11.09	10.25	9.56	8.96	8.42	7.97	7.43	6.72
2013	16.25	12.62	11.26	10.27	9.59	9.03	8.57	8.04	7.54	6.84

Table 8: 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 9: 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			

is different from the baseline (W). 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 SC gap is smaller than the W gap, 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 the W sample. On the other hand, in many of the later years, the SC gap is much greater.

Table 10: Conventional Measures of The Gender Gap (with selection correction)

Year	Mean (1)	10th (2)	25th (3)	50th (4)	75th (5)	90th (6)
1976	0.294	0.200	0.274	0.318	0.335	0.344
1977	0.282	0.136	0.239	0.315	0.354	0.376
1978	0.357	0.210	0.315	0.391	0.425	0.432
1979	0.389	0.249	0.343	0.427	0.459	0.460
1980	0.344	0.198	0.299	0.384	0.420	0.422
1981	0.330	0.186	0.280	0.364	0.404	0.412
1982	0.302	0.140	0.239	0.340	0.383	0.397
1983	0.231	0.025	0.141	0.268	0.342	0.375
1984	0.107	-0.118	-0.001	0.134	0.235	0.289
1985	0.294	0.142	0.231	0.325	0.375	0.397
1986	0.273	0.114	0.213	0.308	0.360	0.367
1987	0.314	0.185	0.269	0.354	0.384	0.379
1988	0.233	0.086	0.166	0.255	0.319	0.338
1989	0.189	0.033	0.124	0.216	0.274	0.304
1990	0.239	0.098	0.187	0.266	0.315	0.328
1991	0.171	0.030	0.110	0.192	0.246	0.277
1992	0.216	0.088	0.175	0.242	0.281	0.295
1993	0.071	-0.120	-0.015	0.095	0.179	0.226
1994	0.039	-0.196	-0.078	0.071	0.176	0.228
1995	0.079	-0.131	-0.011	0.113	0.196	0.235
1996	0.262	0.181	0.233	0.280	0.298	0.302
1997	0.218	0.109	0.181	0.237	0.270	0.282
1998	0.044	-0.206	-0.060	0.079	0.173	0.222
1999	0.176	0.040	0.124	0.189	0.236	0.268
2000	0.216	0.116	0.178	0.232	0.262	0.282
2001	0.187	0.092	0.140	0.190	0.225	0.270
2002	0.246	0.167	0.209	0.248	0.276	0.312
2003	0.241	0.142	0.212	0.248	0.277	0.312
2004	0.227	0.139	0.196	0.237	0.262	0.297
2005	0.246	0.175	0.217	0.249	0.274	0.302
2006	0.284	0.230	0.258	0.284	0.299	0.332
2007	0.283	0.241	0.260	0.283	0.298	0.325
2008	0.208	0.141	0.181	0.209	0.236	0.260
2009	0.145	-0.013	0.082	0.168	0.216	0.259
2010	-0.115	-0.627	-0.454	-0.079	0.140	0.339
2011	0.094	-0.059	0.013	0.108	0.181	0.231
2012	0.161	0.010	0.096	0.181	0.237	0.277
2013	0.259	0.186	0.236	0.273	0.292	0.309

Both the sign and magnitude of the SC gap change at certain percentiles. While a positive gap persists at the upper tail of the distribution between men and women, the SC gap is sometimes even negative in

Table 11: Entropy Measures of The Gender Gap (with selection correction)

Year	S_ρ (1)	Theil (2)	$k=0.1$ (3)	$k=0.2$ (4)	$k=0.3$ (5)	$k=0.4$ (6)	$k=0.5$ (7)	$k=0.6$ (8)	$k=0.7$ (9)	$k=0.8$ (10)	$k=0.9$ (11)
1976	6.712	28.898	2.913	5.554	8.150	10.744	13.382	16.116	19.016	22.216	26.213
1977	6.919	33.004	3.415	6.088	8.727	11.391	14.148	17.087	20.362	24.350	30.731
1978	9.400	46.978	4.787	8.424	11.996	15.605	19.361	23.408	27.991	33.695	43.082
1979	10.432	47.648	4.768	8.914	12.954	16.993	21.132	25.489	30.227	35.658	42.916
1980	8.712	39.752	4.421	7.883	11.302	14.754	18.325	22.131	26.371	31.532	39.793
1981	8.074	36.056	3.950	7.040	10.098	13.188	16.383	19.782	23.563	28.160	35.548
1982	7.288	36.214	3.383	6.146	8.850	11.566	14.369	17.349	20.650	24.585	30.444
1983	5.396	26.438	2.254	4.348	6.382	8.407	10.468	12.611	14.892	17.392	20.285
1984	3.706	17.684	1.737	3.155	4.556	5.967	7.418	8.950	10.630	12.620	15.630
1985	6.075	32.740	2.956	5.248	7.517	9.811	12.186	14.717	17.540	20.994	26.605
1986	5.198	23.740	2.491	4.515	6.520	8.543	10.623	12.815	15.214	18.058	22.421
1987	5.700	23.568	2.434	4.638	6.820	9.006	11.224	13.509	15.914	18.554	21.903
1988	3.986	18.056	2.000	3.501	5.000	6.522	8.100	9.783	11.666	14.004	17.997
1989	3.365	17.152	1.554	2.770	3.979	5.202	6.464	7.802	9.284	11.082	13.985
1990	3.603	15.384	1.837	3.201	4.569	5.960	7.403	8.940	10.661	12.805	16.536
1991	2.487	10.892	1.348	2.274	3.209	4.167	5.169	6.250	7.487	9.094	12.128
1992	2.937	13.862	1.311	2.470	3.615	4.765	5.936	7.148	8.436	9.878	11.803
1993	2.233	10.392	1.290	2.142	2.999	3.880	4.807	5.820	6.999	8.567	11.608
1994	2.554	11.670	1.489	2.492	3.491	4.513	5.590	6.770	8.146	9.967	13.400
1995	2.163	9.166	0.998	1.874	2.737	3.602	4.484	5.403	6.387	7.498	8.982
1996	3.227	18.876	1.622	2.798	3.981	5.189	6.443	7.783	9.290	11.191	14.594
1997	2.562	7.738	1.548	2.476	3.424	4.409	5.457	6.614	7.990	9.903	13.936
1998	2.863	15.364	1.754	2.795	3.830	4.898	6.048	7.348	8.936	11.179	15.782
1999	2.134	10.920	1.205	1.940	2.688	3.464	4.287	5.195	6.273	7.759	10.845
2000	2.370	9.346	1.276	2.156	3.044	3.955	4.906	5.932	7.103	8.622	11.484
2001	2.010	6.594	1.298	1.973	2.671	3.406	4.203	5.109	6.231	7.891	11.682
2002	2.798	16.018	1.426	2.396	3.376	4.381	5.434	6.572	7.877	9.584	12.830
2003	2.673	13.836	1.357	2.322	3.294	4.288	5.323	6.432	7.687	9.288	12.212
2004	2.190	8.026	1.096	1.903	2.719	3.552	4.414	5.328	6.345	7.613	9.862
2005	2.554	11.330	1.536	2.473	3.429	4.421	5.473	6.631	8.001	9.890	13.826
2006	3.281	19.990	1.742	2.941	4.152	5.394	6.691	8.091	9.689	11.766	15.682
2007	3.148	12.164	1.585	2.716	3.860	5.031	6.247	7.546	9.007	10.864	14.265
2008	1.886	6.274	1.232	1.895	2.580	3.299	4.074	4.948	6.020	7.581	11.088
2009	1.743	6.658	1.174	1.804	2.452	3.131	3.865	4.697	5.721	7.218	10.570
2011	1.249	5.250	0.815	1.259	1.717	2.197	2.714	3.296	4.007	5.038	7.331
2012	1.581	7.498	0.696	1.220	1.747	2.284	2.839	3.426	4.077	4.879	6.265
2013	2.352	10.162	1.084	1.938	2.797	3.670	4.565	5.504	6.527	7.752	9.753

[1] Note that all entropy measures are multiplied by 100. The original values S_{rho} are normalized to be between 0 and 1. k correspond to varying levels of inequality aversions.

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 in section 2, we emphasize the importance of decision-theoretic framework in summarizing the information at the distributional level. A summary measure like S_ρ, which accounts for all the moments of the distribution as well as inequality/dispersion, is more preferred.

We find that the SC gender gap is larger in the upper tail than in the lower tail of the wage distributions. Regardless of the measure, the SC gender gap evolves very differently compared to the baseline. It is much more cyclical and fluctuates even more in the lower tail of the distribution. These results are summarized in Figure (5) to be compared with the W case in Figure (1).

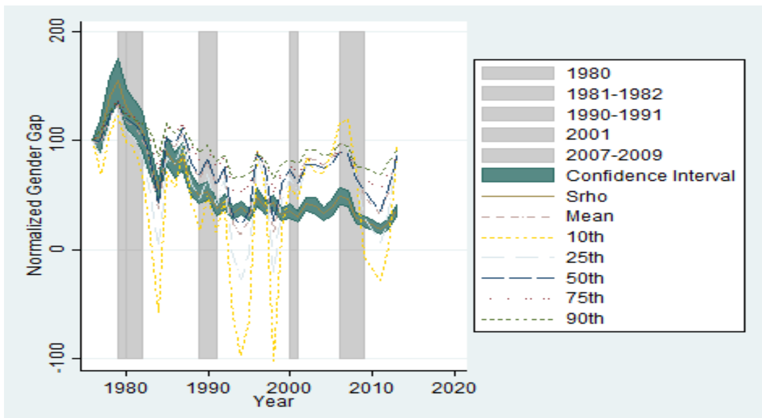


Figure 5: The Trend of SC Gender Gap (Shaded areas correspond to recession periods designated by NBER)

Three-Phase Trend We plot the smoothed trend line for each of the gap measures 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 for W. We now observe 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 the W sample, but with different magnitudes. The differences vary across measures, and can be seen by comparing Table (12) to Table (3). The SC gender gap converges at slower rates in the upper tails, but convergence in the lower tail is more pronounced. The second phase of the trend is even more distinct. 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 great recession, is noteworthy. While there is a lack of convergence in the upper tail, a generally 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 SC gap increased from the beginning of the recession, while W 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). The difference between the SC and W gaps 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

Table 12: 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 (15) as well.

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 full-time 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. There is no stochastic dominance ranking in quite a few periods, making the choice of a summary measure both necessary and sensitive. All entropy measures suggest that, although there was some narrowing trend of the gap in early 1990s, such progress among women as a group 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. Since the patterns are robust to the choice of entropy functions, we report only the S_ρ and Theil measures here (the results using other entropy measures can be found in Table (A.4) in the supplemental material). Subsequently we focus on these two measures.

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.

4.2.3. 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.[32] Dominance relations are then less likely once selection for women is corrected. The impact of addressing selection for *both* men and women requires further attention.

Table (13) summarizes the outcome of SC statistical tests of dominance. Given the relatively large decline in the gender gap during the early years, there is no *statistical* dominance relations in 4 cases

[32] Negative selection implies that extremely high wage earners drop out of 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.

(1977, 1986, 1990, and 1992), for which FSD is observed but at low degree of confidence (less than .90), and no dominance observed 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 informative. It implies that in many of these years, 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.

Table 13: Stochastic Dominance Results with Correction for Selection(Female v.s. Male Wage Distributions)

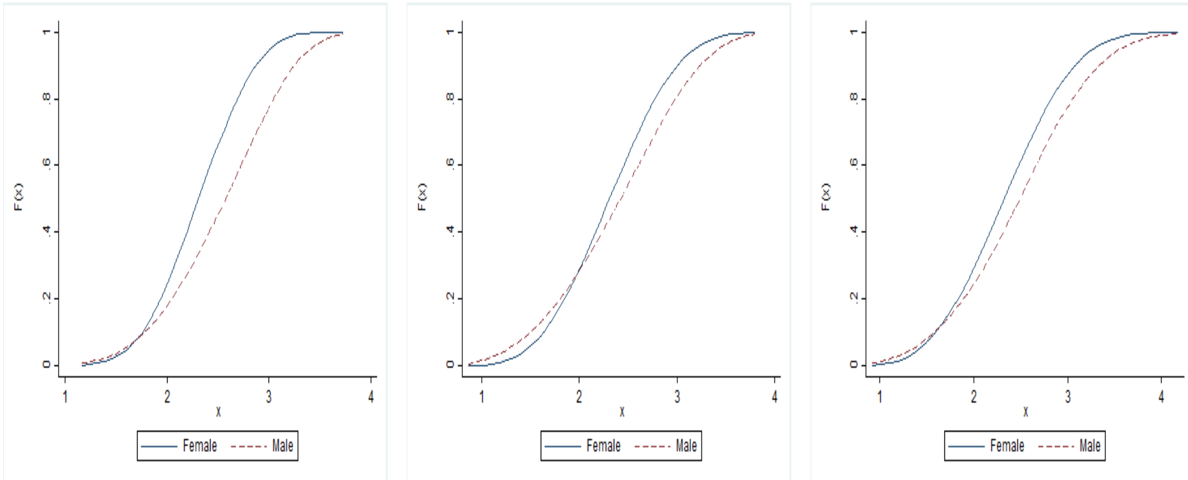
Year	SD	d	$Pr[d \leq 0]$	s	$Pr[s \leq 0]$	Year	SD	d	$Pr[d \leq 0]$	s	$Pr[s \leq 0]$
1976	F	-6.99	1.00	-6.99	1.00	1995	N	59.51	0.00	540.75	0.00
1977	F	-7.74	0.84	-7.74	0.84	1996	F	-6.81	1.00	-6.81	1.00
1978	F	-7.34	1.00	-7.34	1.00	1997	F	-7.16	0.51	-14.06	0.51
1979	F	-8.02	1.00	-11.21	1.00	1998	N	84.45	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. Dominance ranking during the most recent recession is not likely.

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 (6) and the results for other years are available in the supplemental material.

These graphs are illuminating. For all the non-dominated 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 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



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

Figure 6: CDF Comparisons of Female (Selection Corrected) and Male Wage Distributions For Select Years

distributions. *A much 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" interpersonal comparisons of well being. In such situations, it is more than usually important to be explicit about the properties of any evaluation function employed to characterize the gap.

5. Further Explorations by Education and Race

We explore the gender gap by education and race.[33,34] Given constraints of space, we first summarize the common findings in these further explorations and contrast them to our SC 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, with detail estimates 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 SC results. Specifically, we again find slower convergence in the gap or even lack of convergence in some cases. The aforementioned three-phase trend of the gap is observed among some educational or ethnic groups.

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

[34]The subgroup analysis is based on the conditional distributions identified by examining the unconditional distribution (the simulated data) for each subgroup. We do not re-estimate our model for each subgroup. This also ensures that the subgroup and population models are indeed consistent with each other and can be reconciled.

5.1. By Education

We repeat our analysis for four different education groups (below high school education, high school, some college, and college and above). The actual W estimates are reported in Tables (E.6)-(E.9) in the supplemental material, and the SC results are in (E.10)-(E.13). The gender gap among the least educated workers is larger than the gender gap in the rest 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. We highlight the differences in the implied long-run changes for each educational group characterized by both entropy measures in Table (14), and the corresponding graphs are in Figure (E.3) in the supplemental material.

The W results are in Panel A of Table (14). Regardless of educational groups, we find some notable convergence in the gap before the mid-1990s but a modest one afterward. 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 regarding the progress of the least-educated women is very similar to what is found in Blau (1998).

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. Our SC results (Panel B of Table 14) indicate the answer is: No. Even after addressing the selection, the gap among the least educated workers exhibits a slower convergence in the first phase or no progress during the most recent recession, compared to other educational groups, and ever increased during the second phase.[35] And "the deteriorating relative economic position of less educated women" over this period could be well underestimated in Blau (1998).

Turning to SD tests (W results are reported in Tables (A.5)-(A.8) in the supplemental material, and the SC results in (A.9)-(A.12), we again find 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 since we observe dominance relations.[36] But with the SC results, we find that despite some convergence, dominance relations exist in nearly all years for individuals with more than college education,

[35] In the most recent recession, 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.

[36] The W results for individuals with more than college education, some college, or high school education are very similar to the full-sample W results, with generally statistically significant FSD rankings. Further, although generally statistically insignificant, we also observe dominance relations among the least educated individuals.

Table 14: Implied Long-Run Annual Changes in the Gender Gap by Education

Period	S_ρ Less than High School (1)	S_ρ High School (2)	S_ρ Some College (3)	S_ρ College & Above (4)	Theil Less than High School (5)	Theil High School (6)	Theil Some College (7)	Theil 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

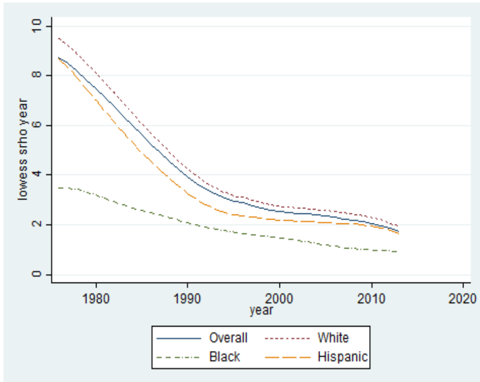
[1] These values are long-run compound annual change rates implied by the initial and the last smoothed values of each period.

while for the rest of the population, the occurrence of such relations decreased drastically during the period of convergence, but increased again in the later years (beyond 1994). This result implies that women with college education fare worse than their male counterparts, while women with less education may not. This latter 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.

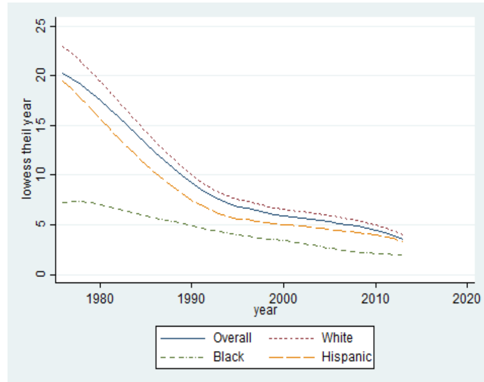
5.2. By Race

The gap by race is reported in Tables (A.13)-(A.15) in the supplemental material, and those for SC in (A.16)-(A.18). 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. It is thus not necessarily surprising that 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 (7)).



(a) S_ρ



(b) Theil

Figure 7: Comparison of Smoothed Trend of The Gender Gap with Selection Correction By race (Excluding 2010)

Turning to our SD results without addressing selection (the W results by race are reported in Tables (E.14)-(E.16) in the supplemental material, and the SD results are in (E.17)-(E.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, for 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. Extension: 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. However, wages may not necessarily be a good measure of women's actual well-being for those who do not work, and the comparison of wage offers does not fully serve our 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. It is then important to take into account the non-market value of time for those who do not work in measuring the gender gap. In economic theory, the actual monetary value of not working (or the best alternative to working full-time) is captured by reservation wages. An interesting yet useful comparison would be based on an alternative wage distribution for men and women, replacing the wage offers with *reservation wages* for those who do not work full-time. Recall that the selection mechanism can be thought of as follows

$$S = I(\ln(w) \geq Y^{\text{reservation wages}}) \tag{14}$$

The alternative wage distribution is thus equivalent to the distribution of $\max(\ln(w), Y^{\text{reservation wages}})$. Our quantile approach allows such analysis. With further structure in the selection equation, we can recover the distribution of reservation wages given unemployment. Specifically, we further impose an additive structure of reservation wages given by $R(z) + \eta$, and the labor force participation is based on the comparison of wage offers and reservation wages:

$$S = I(\ln(w) \geq R(z) + \eta) \tag{15}$$

As noted in Arellano and Bonhomme (2017), this is equivalent to

$$S = I(v \leq F_{\eta - \ln(w)|Z}(-R(z)|z) \quad v|x \sim Uniform(0,1) \tag{16}$$

where $v \equiv F_{\eta - \ln(w)|Z}(-R(z)|z)$ is the standard uniform. Therefore, all the assumptions for qunatile selection models in Section 4.1.2 are met, and the wage function, $g(x, u)$, is identified. Given $g(x, u)$, we can also identify $R(z)$.

In practice, we assume a linear index for $R(z) = z'\gamma$. For a given quantile, τ, (3) becomes

$$\begin{aligned} S &= I(x'\beta_\tau \geq z'\gamma + \eta) \\ &= I(z'\theta \geq \eta) \end{aligned} \tag{17}$$

the second equality is due to $x \in z$. Once quantile function is identified, $x'\beta_\tau$, we can estimate reservation wages, $z'\gamma$, via propensity score equation $\Phi(z'\theta)$. This involves a three-step procedure. (1) For every individual with $X = x$, we simulate the complete distribution of potential wages by computing $\ln(w) = x\widehat{\beta}_\tau$ for $\tau = 2, \ldots, 98$ and (2) estimate $z'\widehat{\theta}$, the linear index from the probit model. (3) Reservation wage conditional on non-participation status is identified by $x\widehat{\beta}_\tau - z'\widehat{\theta}$ for $\tau = 2, \ldots, 98$ given $S = 0$. Potential wage conditional on participation status is given by $x\widehat{\beta}_\tau$ for $\tau = 2, \ldots, 98$ given $S = 1$.[37]

With these estimates, various gender gap measures and SD tests are presented in Tables (15) and (16). 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 (16). We observe even fewer instances of either first- or second-order dominance relations.

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

Table 15: 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 16: Stochastic Dominance Tests between the Mixed Distributions of Market and Non-Market Values

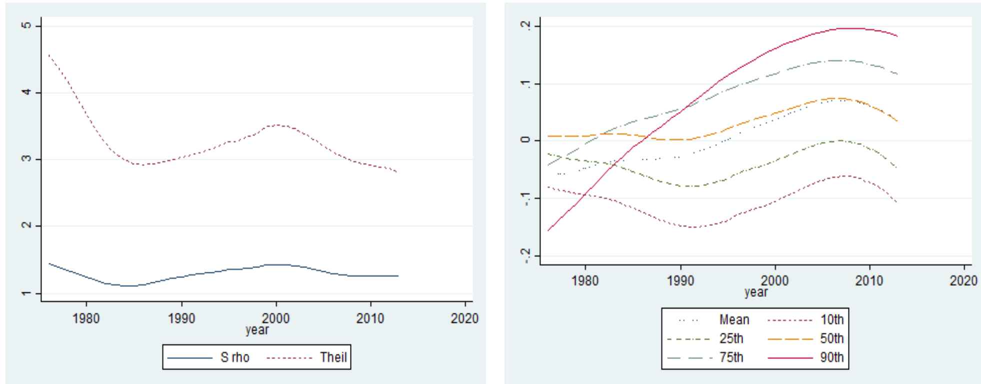
Year	SD	d	s	Year	SD	d	s	Year	SD	d	s
1976	FSD	**-6.28**	**-6.28**	1989	SSD	51.35	**-12**	2002	No	20.87	117.13
1977	FSD	**-8.16**	**-8.16**	1990	No	44.54	349.9	2003	No	22.37	135.35
1978	No	42.3	32.05	1991	SSD	45.4	**-7.72**	2004	No	26.67	170.83
1979	No	29.4	87.12	1992	No	41.56	293.46	2005	No	17.42	96.57
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.

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 (8). 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 recent 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.

7. Implications

In this section, we use the estimated wage distributions for both men and women to further assess and challenge a variety of related assumptions, hypotheses, and findings in some influential works 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.



(a) Entropy Measures (b) Conventional Measures

Figure 8: Comparison of Smoothed Trend of The Gender Gap using Mixed Distributions (excluding 2010)

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

Using our results, we first examine (1) the relationship between selection bias and employment rates; (2) the relationship between the gender gap in employment rates and the gender wage gap; (3) and the role of selection in explaining the observed relationship in (2).

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 *negative* 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 gap. 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 gender gap increases (because more low-wage women enter the labor force). 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 so will the role of selection in explaining the relationship.

(a) Female (b) Male

Figure 9: Percentage of Full-time Female Workers and the Estimated ρ (measuring selection)

Using our estimates of selection parameter, ρ, we find results in contrast to common wisdom. First, we find varying relationship between employment rates and the direction and magnitude of the selection across gender and over time. This pattern can be visualized in Figure (9). The solid line is the full-time employment rates among women, and the dashed line is the estimated ρ. For women, there is negative selection when employment rates are extremely low, while there is positive selection with relatively higher employment. In the presence of negative selection, the magnitude of the selection bias becomes smaller as employment rates rise, although the relationship is not necessarily monotonic.[38] In the presence of positive selection, the magnitude of the selection parameter (positive selection) fluctuates closely (rather than decreases) with the employment rates for both men and women. When considering both positive and negative selection together, we fail to find any systematic relationship between employment rates and the *magnitude* of selection bias among women (or men). 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).

Table 17: 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

Second, we compute and report the correlation between employment gaps and wage gaps in Table (17). 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.

Finally, in the presence of positive selection, taking into account selection can explain away most of the correlation between employment and wage gaps, in line with Olivetti and Petrongolo (2008); 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.

[38]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.

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.

Table 18: 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.

Our formal test of this assumption is presented in Table (18), 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 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.. It is, however, worth noting that when testing the monotonicity assumption in Blundell et al. (2007), we maintain all the assumptions of the model (such as the specifications of the copula and quantile models); our results should be considered a joint test of the validity of both sets of assumptions.

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.

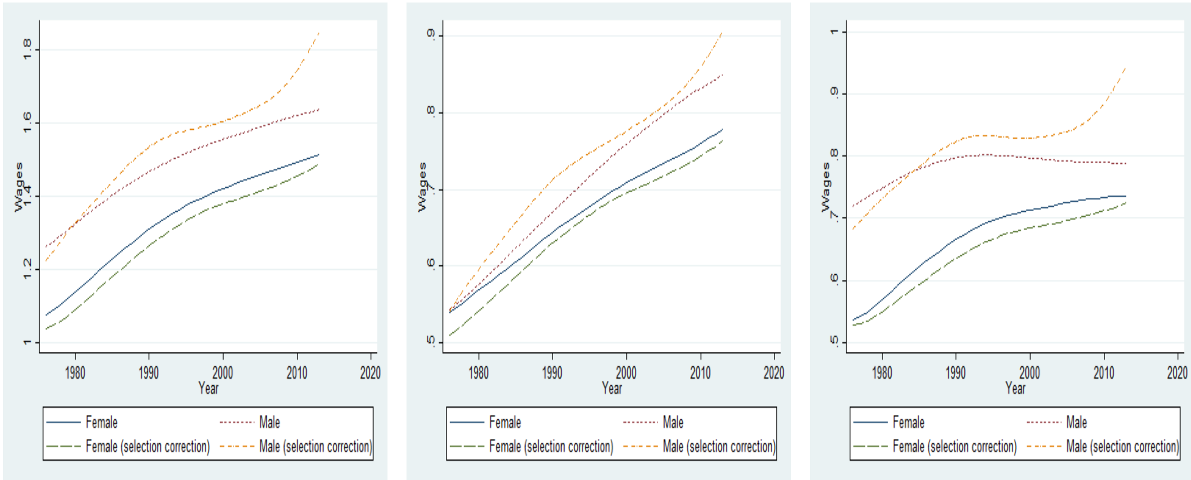
However, these findings are based on *observed* wage inequality, (without considering individuals who do not work), instead of the underlying wage inequality, upon which Mulligan and Rubinstein (2008)'s explanation is built. These two quantities could be drastically different. Not only could addressing the selection and estimating the distributions of potential wages potentially invalidate Mulligan and Rubinstein (2008)'s explanation. It may also challenge the influential results about the patterns in the within-group inequality in the existing literature (e.g., Juhn et al., 1993; Autor et al., 2008).

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 (10): 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. We first discuss the SC results, contrasted with the W results, and then the sources behind the discrepancies.

The SC vs. W Results Without considering non-working individuals, our findings are broadly consistent with Mulligan and Rubinstein (2008) and Autor et al. (2008).[39] When selection is considered, for women, while the general patterns of increasing trend for these measures continue to hold, 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. For men, 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 W results. 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.

Sources behind the Discrepancies between W and SC results Further analysis indicates important differences between our results and those in Mulligan and Rubinstein (2008) and Autor et al. (2008),

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



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

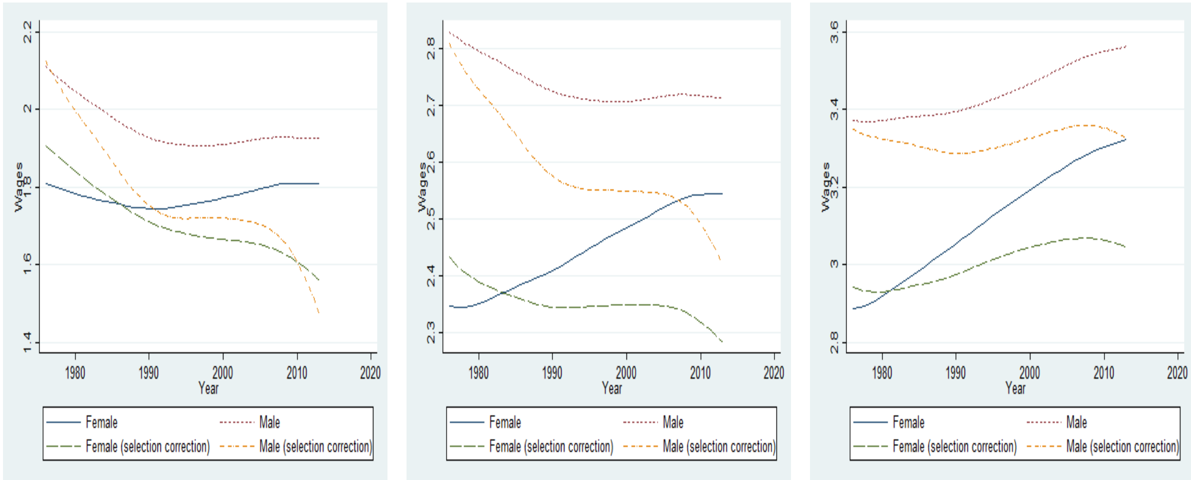
Figure 10: Comparison of Smoothed Trend of Various Inequality Measures

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 (11). It can be seen that the reason for the increase in the *observed* overall inequality (90/10 and 90/50 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 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).

In sum, the implied evolution of wage inequality within and between gender supports a theoretical explanation like Mulligan and Rubinstein (2008), while the patterns for men and the sources behind these patterns found here 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.[40]

[40] 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.[41] 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



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

Figure 11: Comparison of Smoothed Trend of The Distributions for both Men and Women with and without Selection Correction

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 focuses on 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

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).

re-weighting by propensity scores. We adopt the first approach because we are able to estimate the (*true*) conditional quantile selection regressions.[42]

Machado and Mata (2005) is among the first to estimate quantiles to recover the counterfactual distribution, and Chernozhukov et al. (2013) discuss the corresponding inferential theory.[43] 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 (D.2), 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.19)-(A.22) in the supplemental material. We summarize three main findings here.

Structural vs Composition Effects 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.

Composition Effects: W vs. SC 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

[42] 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.

[43] 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.

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).[44] However, when controlling for selection, we find that the effects of such progress on the gender gap 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.

Structural Effects: W vs. SC 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.24)-(A.26) in the supplemental material, we reach some even stronger conclusions and uncover some masked by examining only the gap at select quantiles. We summarize these findings below.

SD Structural Effects: W vs. SC 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.

SD Compositional Effects: W vs. SC 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. An implication is that the *distribution* of women's

[44] 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).

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. [45]

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.

[45] 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 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.

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 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 to selection is necessarily premised on a number of assumptions. For example, presence of young children, a typical choice as an exclusion restriction for selection in the literature, is also used here. While controversial, we provide a more rigorous statistical test of the validity of the exclusion restriction in our context and show that the assumptions are met for the sample of women. Identification is also aided by Copula functions. However, our semiparametric approach preserves quite a bit of flexibility and achieves some efficiency in estimation, while still being more robust than completely parametric approaches in selection models. We find specification errors for the sample of full-time workers are within a small neighborhood. Further robustness checks using alternative, Gaussian Copula (another low-dimension Copula) show nearly identical estimates of the distributions and dependence parameters. While more robustness results are warranted, these results do increase the confidence in our analysis.

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, 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.

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

Arellano, M. and S. Bonhomme. 2017. "Quantile Selection Models: with an Application to Understanding Changes in Wage Inequality." *Econometrica* 85:1–28.

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.

Bitler, M., J. Gelbach, and H. Hoynes. 2008. "Distrbituional Impacts of the Self-Sufficiency Project." *Journal of Public Economics* 92:748–765.

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. 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?" *American Economic Review* 103:251–256.

Blau, F.D. and L.M. Kahn. forthcoming. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature* .

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* .

Bourguignon, F. 1979. "Decomposable Income Inequality Measures." *Econometrica* 47:901–20.

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.

Dalton, H. 1920. "The Measurement of the Inequality of Incomes." *Economic Journal* 30:348–361.

Dehejia, R. H. 2005. "Program evaluation as a decision problem." *Journal of Econometrics* 125:141–173.

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.

Flood, Sarah, Miriam King, Steven Ruggles, and J. Robert Warren. 2017. "ntegrated Public Use Microdata Series, Current Population Survey: Version 5.0. [dataset]." Technical report, University of Minnesota.

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

Frandsen, B. and L.J. Lefgren. forthcoming. "Tesing Rank Similarity." *Review of Economics and Statistics* .

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. and J. Smith. 1998. *Econometrics in the 20th Century: The Ragnar Frisch Centenary*, chapter Evaluating the welfare state. Cambridge University Press.

Heckman, J.J., J. Smith, and N. Clements. 1997. "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogneity in Program Impacts." *Review of Economic Studies* 64:487–535.

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.

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.

Lee, Lung-Fei. 1983. "Generalized Econometric Models with Selectivity." *Econometrica* 51:507–512.

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. 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* .

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.

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.

Shorrocks, A.F. 1980. "The Class of Additvely Decomposable Inequality Measures." *Econometrica* 48:613–25.

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.