The Gender Gap Between Earnings Distributions

Esfandiar Maasoumi
Emory University

Le Wang
University of Alabama, & IZA
October 2014

Abstract

This paper provides an examination of the notion of "distance" and ordering between the whole distributions of earnings for men and women, and some related counterfactuals. We investigate aggregative measures and their statistical implementation and evaluation. We provide a synthesis of the decision-theoretic basis of measures for such things as the "gender gap". "Strong" comparison/ordering of distributions of outcomes requires subjective evaluation functions reflecting decision maker's preferences. Examples include popular notions of the "gender gap" based on means or medians. These are seen to be supported by rather strong subjective evaluation of heterogenous outcomes at different quantiles. We examine a more flexible and well supported distribution metric based on entropies. We then examine complementary (weak) uniform ranking of distributed outcomes over large classes of preferences by such methods and tests as stochastic dominance. The paper's primary purpose is to provide a comprehensive application of these two interrelated approaches based on the most popular data and models in the existing literature for the US labor market for the last several decades. Selection to the labor force is a major issue and requires new techniques for its treatment at the entire distribution level, compared to existing approaches to the conditional mean and median. We adopt a new quantile-copula approach to modeling the participation decision and find selection is consequential for a deeper understanding of the "gap" and its movement over the economic cycle. We also provide decomposition of the gap to market pricing effects and human characteristic effects based on counterfactual distributions derived from the conditional quantiles. We highlight the importance of education by a comparative analysis of different educational groups.

—

† We would like to specially thank Professor James Heckman (the Editor) for many instructive discussions and guidance. We are grateful to four knowledgeable referees for their helpful and constructive comments and suggestions. We thank Stephane Bonhomme for several inspiring discussions of the findings. Part of the project was fundertaken 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 seminar participants including at Aarhus, Infometrics Institute, Georgia State, Bentley University, Union College, Emory University, Jinan University, Kansas State, Alabama, New Hampshire and Southern Economic Association for their helpful comments.

Contents

1. Introduction

The first question we examine is: What is meant by "the gap"? We consider the decision theoretic basis of different functions that may define the gap, and examine the robustness of current findings based on popular notions of the gender gap, such as means and medians, as well as a metric entropy measure that we highlight. A very large "measurement literature" on the gender gap examines the issues and existing challenges, and sheds light on potential policy issues; for example, Blau and Kahn (1997), Blau and Kahn (2006), and Goldin (2014), to name but only a few. Many researchers have noted that "the gender gap" has decreased over time, especially in the 1980s and early 1990s. However, the narrowing trend has slowed down after mid-1990s. Despite a slower rate of decline, this literature suggests an encouraging trend: women are catching up with men. This sentiment is captured by the title of the recent Ely lecture delivered by Claudia Goldin: A Grand Gender Convergence. The "analysis" side of the literature focuses on understanding the issues and problems suggested by the measurements. This side of the literature typically conducts decomposition/counterfactual analysis, dating back to Oaxaca (1973) and his subsequent works such as Oaxaca and Ransom (1994). Recent literature has developed various alternative decomposition methods and further extended such analysis to the nonparametric settings (e.g., Frolich 2007; Mora 2008; Nopo 2008) and the distributional level (e.g., DiNardo et al. 1996; Blau and Kahn 1994; Albrecht et al. 2009; Fortin et al. 2011) This type of analysis often decomposes the "gender gap" into two components: One is attributed to differences in human characteristics (composition effects), and another to market wage structure, the returns to these characteristics (structural effects).

The second question we address is accounting for selection at the whole distributional level. A quantile-copula approach is examined that is recently proposed by Arrelano and Bonhomme (2012). This approach jointly models the wages and the participation decision. We derive our various distributions from these conditional quantiles both to measure the gap and to provide decompositions by means of counterfactual distributions, corrected for selection and uncorrected. Although we present our findings for a number of educational groups, this paper is focused on the gender gap, and not with identification of individual covariate effects and factors.

The concept of "the gap" between two outcome distributions, observed or counterfactual, requires careful consideration to avoid implicit subjectivity. Different measures may reflect radically different aspects of a distribution of outcomes, the weights attached to different parts of the distribution, and inevitably difficult "interpersonal" comparisons of well being. The choice of measure is also important when examining the evolution of the gender gap, since the rate and pattern of changes may vary across different measures, leading to different conclusions. The timing of temporal deviations from the long-run trend can also vary across measures, with conflicting indications of the relation between business cycles and the gender gap. The traditional analysis based on the mean or median, or indeed any single quantile is supported by rather narrow evaluation functions. More flexible and robust functions of earnings distributions are available and deserve examination. This "decision theoretic" view, while well established since at least the 60s and 70s, is only recently gaining in both theoretical and important new econometric tool development in this area. Our empirical examination reveals which aspects of the gender gap are robust to the definition of it, and which are not.[1]

It may not be fully evident that interest in the gender gap is fundamentally a concern with "equity". There is a decision theoretic foundation available for this kind of analysis. If "dispersion" matters, then it matters within groups and between them. There is a rich literature on inequality measures and entropic functions that sheds light on properties of various functionals of a distribution. Every functional representation of a distribution, including the mean and median, is equivalent to a weighting scheme for different members of that population, as well as often implicit assumptions on inter-personal comparisons of well-being.

We examine a particular distance functional which is also subjective, but is a proper metric, as required for measurement of distances (not divergences) between whole distributions, and we expose its decision theoretic properties. It is aggregative of "the gap" and we point out when it agrees with existing views of the gender gap based on the mean or median. The distributional measure of the gender gap is based on the normalized Bhattacharay-Matusita-Hellinger entropy proposed by Granger et al. (2004). One important feature of this measure is its ability to summarize the distance between two *entire* distributions, in both univariate and multidomensional cases. The preference (weighting) function underlying the family of measures considered here is known to possess many desirable properties. The particular member examined here stands out in the class of "ideal measures" of relative entropies (or "inequality measures") by virtue of

[1]See Carneiro et al. (2001). By any objective measure, this type of methodology is still largely absent in empirical labor economics and from public and professional discussion. Important new contributions that we cite in this paper are exceptions, some as yet unpublished.

being the only metric measure satisfying global triangularity. In contrast, the mean will have the same valuation for a million dollar gap between a pair of CEOs as for one dollar of gap between a million low wage individuals. The median is robust to many changes in the tails, so not particularly suited for public policy discussions on such topics as the top "the 1%", or treatment programs designed to address poor employment prospects or alleviate poverty.

In should be acknowledged, however, that there is no universally agreed cardinal measure. All complete rankings are cardinal and subjective, albeit to various degrees. The natural alternative is to examine existence of uniform (weak) ordering between distributed outcomes that are robust to very large sets of decision functions. This requires stochastic (or prospect) dominance rankings and statistical tests for them. If men's earnings first order dominate women's, we can use the mean to characterize it, or add up all the differences for all the observed individuals, or single out any favorite quantile! If the distribution functions cross, there is no first order dominance and the choice of more restrictive evaluative functionals must be made explicit. The mean of the distribution, or any quantile, would appear to be rather radical choices in this situation. Heterogeneity in outcomes is real and it matters.

We employ statistical SD tests that have been previously developed and analysed by us, and have been widely used to analyze inequality and poverty issues and financial outcomes, but (to our surprise) not to analyze the gender earnings gap.

A major concern with empirical analyses of the gender gap is the selection issue. Women have relatively low rates of employment and labor force participation, and we do not observe wages for nonworkers. To the extent that non-working women systematically differ from working women, any measure or analysis of the gender gap could be biased. This type of selection is well known, dating back to Heckman (1974), and special attention has been paid to it in nearly all research that involves studying women's labor market outcomes. However, in the gender gap literature more welcome attention is being paid to this issue, albeit mostly in the context of conditional means or specific quantiles (e.g., Mulligan and Rubinstein, 2008; Olivetti and Petrongolo, 2008).

Recent extension of methods to deal with selection for quantiles and whole distributions is an important development. Labor force participation decisions could be part of the explanation for the observed gender gap. A joint model/distribution for the participation decision and earnings is required for the analysis of the gender gap between entire earnings distributions. We demonstrate an approach proposed by Arellano and Bonhomme (2012) to this problem which is less restrictive than the few alternatives available to date. It relies on copulas, but leaves the marginal distributions to be determined however one wishes.

While comparing the distributions of wage *offers* is certainly interesting and informative, for women who do not work, the wage offers are not necessarily the actual level of "wages". From a standard labor supply model, these individuals derive value from not working, and such value is captured by their reservation wages. It would thus be interesting to take this into account in our comparisons. We provide an additional concept of the gender gap in this paper. This new concept replaces women's wage distribution with a mixed distribution of women's wages conditional on employment and women's *reservation wages* conditional on non-employment. The Arellano and Bonhomme (2012) approach which we adopt for addressing selection provides a useful structure to recover the reservation wages and its distribution using the potential wage offers and the selection mechanism. We believe that this provides a further set of new results that are potentially interesting and policy relevant for understanding women's well-being, which has not been previously discussed and explored in the literature.

The final section of this paper presents our decompositions that shed light on human capital and market return components, with and without correcting for selection into the labor market. Our methodology allows derivation of required counterfactual distributions from the quantiles. This contrasts with our other related work in which this is achieved by inverse probability weighting methods which are not as yet suited to accounting for selection.

A number of important and currently topical extensions of our methods are described in the concluding section. These include multidimensional analysis, the race gap, the inequality, and program evaluation.

The rest of the paper is organized as follows. A brief summary of findings is given next, followed by a section on methodology which describes a decision theoretic basis for our entropy measure, as well as dominance testing. Data is next described and the empirical findings are given, first without and then with accounting for selection. Concluding remarks are given in the final section.

2. Summary of findings

Our findings are based on the Current Population Survey (CPS) data, 1976 – 2013. They are based on full implementation of rigorous statistical procedures and to a degree of statistical confidence. Some of our findings may be conveniently summarized as follows:

1. The discrepancies between various conventional measures of the gender gap highlight the need for a synthesis. These results facilitate comparisons to other findings in the literature that do not address the selection issue.

- All measures show a consistently declining long-run trend of the gender gap, but the timing of temporal deviations from the long-run can be measure-specific. This too presents a conundrum when assessing the association of business cycles and the gender gap. The metric entropy measure indicates an aggregation approach that is less volatile and indicative of more stable long run trends.
- Without correction for selection or controlling for education, the overall *observed* earnings distributions are generally first-order stochastically ordered to a degree of statistical confidence throughout the sample period, implying that men have generally fared better. This

conclusion is robust to a wide class of increasing Evaluative Functions (EFs). Without selection, any measure would be adequate for "ordering" outcomes, but of course would differ in magnitude.

- Based on updated conventional measures of the gender gap we find that the *observed* gender gap dropped notably before 1990s, but the convergence has slowed down since. This result is consistent with the pattern of the "grand gender convergence" found in the literature.
- The entropy gap is a function of all the quantiles and moments of the distribution and consistent with flexible decision and preference functions. Uncorrected for selection, it indicates greater convergence of women and men's earnings than is found in the literature. Traditional summary measures underestimate the declining trend of the *observed* gender gap in the U.S. The entropy gap narrowed at an average annual rate of 6% during the period 1976-2013, while the largest annual rate recorded by the conventional measures is 2.7% (at the median).

2. A second set of results highlight the importance of addressing selection.

- Selection matters and varies over time, from negative to positive (consistent with Mulligan and Rubinstein (2008)). The *population* (selection corrected) gender gap is overestimated in early years while underestimated in more recent years. Women's selection into employment accounts for the majority of the observed narrowing trend of the gender gap. Specifically all the conventional measures fluctuate around their starting levels before the most recent recession, which suggest that the *population* gap may have not shrunk at all! Our selection corrected mean and median gap are strikingly similar to Mulligan and Rubinstein (2008) and Olivetti and Petrongolo (2008) using completely different approaches. The entropy measure indicates that the gender gap declined before 1990s, but remained stagnant throughout the 1990s and early 2000s.
- The "convergence" trend may have reversed since the most recent recession. The gender gap increased notably since 2007, and declined slightly in 2013. Comparison of the results before and after addressing the selection indicates that the convergence in the *observed* gender gap (typically found in the literature) may be explained by the selection of women with low-wage offers out of employment during the recession.
- We find several cases in early years where no statistically meaningful dominance order holds. In these cases, the inference that men fare "uniformly" better than women can only be supported by a narrower class of preference functions that are more than merely increasing and concave. Orderings are not meaningfully "uniform" and are highly subjective.
- Our comparative analysis of different educational groups indicates that women with college education do not necessarily fare worse than their male counterparts, while women with less education do. This highlights the importance of education as a major human capital element. This also 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.

- A new alternative gender gap measure (based on the mixed distribution of wage offers and reservation wages conditional on employment status) implies that women's relative well-being may have deteriorated over time, also in stark contrast to the results without addressing the selection.

3. Using the proposed techniques and addressing selection at the distributional level provides insights on some other questions.

- We are able to analyze the relationship between full-time employment rates and the pattern of selection among women. We find that selection does not disappear as women participate more in full-time employment, as conjectured in Smith and Ward (1989). Before 2001, selection increases from negative to positive as women's full-time employment increases. After 2001, the magnitude of the positive selection continued to increase while full-time employment of women began to decline.
- Within-group inequality in the distribution of potential wages has increased over time, which is an explanation for the transition in the selection pattern proposed in Mulligan and Rubinstein (2008). This result is notable because the latter analyzed within-group inequality for *working women* only.
- The ability to obtain distributions of potential wages for the whole population allowed us to formally test first order dominance relation which is assumed in Blundell et al. (2007). We fail to find evidence supporting that assumption in more than half of the sample period, indicating that working women do not necessarily perform better than non-working women.
- The "structural effects" are generally greater in magnitude than the "composition effects". The dominance tests, after addressing the selection, imply that policies aimed at the wage structure or improving human capital characteristics for women could *uniformly* improve their earnings distribution.

3. Methodology

3.1. Basic Notation

To begin, let $\ln(w^f)$ and $\ln(w^m)$ denote the log of earnings for females and males, respectively. We observe a random sample of $N = N_1 + N_2$ individuals. $\{\ln(w^f)\}_{i=1}^{N_1}$ is a vector of N_1 observations of $\ln(w^f)$ (denoted by $D_i = 1$); similarly, $\{\ln(w_i^m)\}_{i=1}^{N_2}$ is a vector of N_0 observations of $\ln(w^m)$ (denoted by $D_i = 0$). Let $F_1(y) \equiv Pr[\ln(w^f) \leq y]$ represent the cumulative density function (CDF) of $\ln(w^f)$ (i.e. the log of earnings for females) and $f_1(y)$ the corresponding probability density function (PDF); $F_2(y)$ and $f_2(y)$ are similarly defined for $\ln(w^m)$ (i.e. the log of earnings for males).

3.2. Decision Theoretic Foundations of Various Measures of The Gap

Usually, the gender gap is defined as the difference in certain functions of the earnings distributions. For example, average gender gap is the difference in the means of the earnings distributions ($\mathbb{E}[\ln(w_i^m)] - \mathbb{E}[\ln(w_i^f)]$) (thus the first moment of the earnings distribution). The gender gap at a τ^{th} quantile is $Q_\tau(\ln(w_i^m)) - Q_\tau(\ln(w_i^f))$, where the τ^{th} quantile of F_1 (the CDF for women's wage distribution) is given by the smallest value $Q_\tau(\ln(w^f))$ such that $F_2((\ln(w_i^m)) = \tau$; $Q_\tau(\ln(w^m))$ is similarly defined for F_2 (the CDF for men's wage distribution).

Recent work on identification of earnings and counterfactual distributions pays greater attention to suitable functions of the underlying distributions; see Chernozhukov et al. (2013) and Firpo et al. (2007). The Gender Gap is an example of such functionals. Following Carneiro et al. (2001), we consider decision theoretic implications of various gap functions.

Several axioms may embody what we consider as desirable in any measure/index and motivate its use. A measure of distance between a pair of distributions for the random variable X may be required to satisfy the following"ideal" properties.

1. It is well defined for both continuous and discrete variables.
2. It is *normalized* to $[0, 1]$.
3. It is well defined and applicable when X is be multidimensional.
4. It is *metric*, that is, it is a true measure of "distance" and not just of divergence.
5. The measure is *invariant* under continuous and strictly increasing transformations $\psi(\cdot)$, such as logarithmic.

We propose a normalization of the Bhattacharya-Matusita-Hellinger measure given by

$$\begin{aligned} S_\rho &= \frac{1}{2}\int_{-\infty}^{\infty} \left(f_1^{1/2} - f_2^{1/2}\right)^2 dx \qquad (1) \\ &= \frac{1}{2}\int \left[1 - \frac{f_2^{1/2}}{f_1^{1/2}}\right]^2 dF_1 \end{aligned}$$

where f_1 and f_2 are the two densities. The second expression is in moment form and can be used for fast, but inaccurate computation. Importantly, $\mathrm{H}_0 : (f_1 = f_2) \Longleftrightarrow S_\rho = 0$, otherwise (under H_1) $S_\rho > 0$. Power and *consistency* of the tests based on consistent estimates of S_ρ arise from this property.[2]

The proposed measure satisfies all the desirable properties stated above. Note that log of earnings is an increasing function, and our measure of the gap is invariant to the choice of levels or logs of wages. A more subtle advantage of this and other entropic measures is that they are defined over the space of *distributions*. This makes them "dimension-less", able to seamlessly

[2]Integrated squared norm ($L2$) also shares many of these properties, but it is not normalized and thought to be more sensitive to "inliers and outliers' (Hart, 1997).

accommodate univariate and multivariate contexts. This is not so for measures defined over the variable space or based on quantiles. Economists have been increasingly aware that evaluation of individual well-being is inevitably a multi-attribute exercise (Maasoumi, 1999). This feature is advantageous when *multidimensional* versions of gender gap, inequality, mobility and poverty are examined which go beyond wages.

Our proposed measure is supported by additional decision theoretic properties, to which we now turn. Dalton (1920) is credited with the earliest statement of a formal correspondence between functions of distributions, like inequality measures, and "Social Welfare Functions" (SWFs). Atkinson and Kolm, made this connection much more precise. See Kolm (1969) and Atkinson (1970).

Under a set of axioms on policy maker's preferences Yaari (1987) shows that they can be presented by an evaluation function (EF). The form below is an example of such representations that we find revealing here:

$$EF(x) = \int_0^1 p'(\tau)F^{-1}(\tau)d\tau \tag{2}$$

where F is the CDF of earnings, $p(\cdot)$ is the preference function of the policy maker, and $p'(\cdot) \geq 0$ the weight assigned to $F^{-1}(\tau)$, the τ^{th} quantile. EF is a weighted average of "earnings" quantiles. Ranking distributions is equivalent to comparing the evaluation functions. A single quantile (like the median) is especially undesirable when there is heterogeneity and dispersion. Yaari (1988) and others have shown that the Pigou-Dalton Pinciple of Transfers is satisfied if and only if $p''(\cdot) \leq 0$ (Zoli 1999). Evaluation based on a single, or a subset of quantiles is equivalent to assigning zero weights $p'(\tau) = 0$ to all others. Such a weighting scheme is not consistent with any (reasonable) preference functions with heterogenous outcomes. Mean earnings, the most commonly employed measure of the gap, assigns equal weights to each wage earner. This is seldom how policy interventions are targeted or evaluated.[3]

The subjectivity of any aggregation decision is inevitable and a matter of degree. Some guidance is needed on desirable measures of dispersion/inequality, and to reveal the EF underlying our gap measure. For a review of this literature see Maasoumi (1998). Concave and increasing Evaluation Function of a policy maker (represented by Equation 2) is known to be similarly represented as a function of inequality:

$$EF = \mu_x[1 - I(x)] \tag{3}$$

Let x_e denote the "equal equivalent earning"; i.e., the level of earning which if received by everyone would leave total Evaluation (EF) at the same level as for a given earnings distribution.

[3]Reporting a vector of quantiles will leave it to the observer to decide, but also assumes a dollar is the "same" for a billionaire as for a destitute.

Indeed we may take EF $= x_e$. Note that $x_e < \mu_x$ so long as there is any inequality. The larger the difference between x_e and μ_x, the larger $I(x)$. As argued in Atkinson (1970) and Kolm (1969), $I(x)$ is actually a measure of relative inequality, and $[1 - I(x)]$ indicates the degree of loss due to inequality.

It is clear that a decision theoretic comparison of two distributions based on EF is equivalent to a comparison based on inequality measures, adjusted by their means (when they differ). It is equally clear that this approach does not by itself identify a unique inequality index.[4] There are many choices of inequality measures consistent with this approach. One important example is the Gini coefficient and the corresponding welfare function, EF, is equal to the well-known Gini social welfare function (Sen 1974; Aaberge et al. 2013).

A flexible family of "ideal" measures of (relative) inequality is the Generalized Entropy (GE) family of indices:

$$I_\gamma = \frac{1}{\gamma(\gamma+1)} \int_0^\infty (\frac{x}{\mu_x})[(\frac{x}{\mu_x})^\gamma - 1]dF, \gamma \in R \tag{4}$$

where γ is the degree of aversion to relative inequality; and for $\gamma < 2$, the higher its absolute value the greater is the sensitivity to transfers in the lower tail of the distribution. By a monotonic transformation, the Atkinson family of inequality measures is equivalent to GE.

Equalizing transfers between millionaires are (increasingly) less valued than between richer and poorer, or amongst poorer individuals. This family includes Herfindahl's, variance of logarithms, square of the coefficient of variation, and Theil's first and second measures, I_0 and I_1, respectively.

Fundamental axioms of symmetry, continuity, Principle of Transfers, and decomposability, identify GE as the desirable scale invariant family of inequality measures, see Bourguignon (1979) Shorrocks (1980, 1984) . The axiomatic discussion of the properties that identify GE is constructive and reveals degrees of subjectivity in all meaaures of inequality. Further "restrictive properties" are required to justify members of GE. A commonly imposed property is a more stringent additive decomposability, or aggregation consistency. See Shorrocks (1984). If we require that total inequality must increase if inequality increases for every subgroup of a population, and should be invariant to subgroup inequality weights, we reach the conclusion that Theil's second entropy/inequality measure is the ideal measure. This kind of decomposability is very useful for controlling and dealing with heterogeneity of populations, and as a means of unambiguously identifying demographic, geographic and other sources of the gap/inequality and those that are affected by it. Such additive decomposability and aggregation consistency criterion is violated by Gini and variance of logarithms. Absent First or Second order Stochastic Dominance, the class of preference functions that woud support "uniform rankings" are far

[4] To avoid choosing an inequality measure and conduct uniform comparisons is seen to be equivalent to ranking by Lorenz curves, adjusted for possibly unequal means, the so called Generalized Lorenz dominance, which is identical to second order stochastic dominance.

from "consensus" for policy decisions. This is nicely demonstrated in Aaberge et al. (2013) which reveals that third and higher uniform rankings are (effectively) Gini or Generalized Gini rankings! This is so unless we insist on aforementioned additive decomposability properties.

Another desirable property that may be imposed to restrict the class of inequality/entropy/EF functions is "metric-ness". The generalized entropy family are measures of "divergence" between a distribution of earnings and a corresponding population distribution (the uniform distribution when all members receive an equal weight). Divergence is generally not distance. It follows that the divergence between two entropy inequality functions is thus a measure of the "gap" between them. Any member of GE family would do for binary, pairwise evaluation of earnings distributions, but not so when several are needed to be compared, or at different times. There is a single member of the GE family that provides a metric/distance function, the one examined in this paper. A full discussion is given in Maasoumi (1993) and summarized in Granger et al. (2004). For any two density functions f_1 and f_2, the asymmetric (with respect to f_2) k-class entropy *divergence measure* is:

$$I_k(f_2, f_1) = \frac{1}{k-1}\left[\int (f_1^k/f_2^k)dF_2 - 1\right], \quad k \neq 1, \tag{5}$$

such that $\lim_{k\to 1} I_k(\cdot) = I_1(\cdot),$. Once divergence in both directions of f_1 *and* f_2 are averaged, a symmetric measure is obtained which, for $k = 1$, is well known as the Kullback-Leibler measure. These measures fail to be globally *metric* since they violate the triangularity rule for a *distance*.[5]

On the other hand, consider the symmetrized k-class measure at $k = \frac{1}{2}$ as follows:

$$\begin{aligned} I_{1/2} &= I_{1/2}(f_2, f_1) + I_{1/2}(f_1, f_2) = 2M(f_1, f_2) \\ \text{where } M(\cdot) &= \int (f_1^{1/2} - f_2^{1/2})^2 dx \text{ is known as the Matusita-Hellinger } \textit{distance} \end{aligned} \tag{6}$$

$M(\cdot)$ is rather unique among measures of divergence since it satisfies the triangle inequality. The measure of the gender gap examined here,

$$S_\rho = \frac{1}{2}M(\cdot) = \frac{1}{4}I_{1/2}$$

is a normalized $M(\cdot)$ and special member of the GE family. Its unique Evaluative Function is revealed in (Equations for EF) above, when EFs are differenced for two distributions. We wish to emphasize that Quantile-by-quantile measures are informative, and simple aggregates of them should be highly correlated with S_ρ. But a vector of quantiles in front of an observer invites the view that a dollar in the hands of a millionaire is the same as in the hands of a potential participant in a treatment program.

[5]The Kullback-Leibler measure is locally a metric, as shown in Balasubramanian et al. (2014)

Following Granger et al. (2004) and Maasoumi and Racine (2002), we consider a robust nonparametric kernel-based implementation of (1) (The computer code **-srho-** written by the authors in Stata is also available upon request). In our illustrative example below, we use Gaussian kernels and a more robust version of 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.). Interested readers are referred to Li and Racine (2007) for more sophisticated bandwidth selection procedures. 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). The asymptotic distribution of the feasible measure has been derived by Skaug and Tjostheim (1996) and Granger et al. (2004). However, these asymptotic approximations are well known to perform very poorly in almost every case examined. Following Granger et al. (2004) and Maasoumi et al. (forthcoming), we instead employ bootstrap re-sampling procedure based on 299 replications to conduct statistical inference.

3.3. Uniform Ordering: Stochastic Dominance Tests

Any specific definition of the gap provides a "complete" (cardinal) ranking of two wage distributions. When distributions cross (especially at lower tails)[6], such measures of gap may be the only subjective means of ranking outcomes. It is useful then to test whether distributions cross with statistical degrees of confidence. If they satisfy certain conditions, uniform ranking over larger classes of preference functions is feasible. Conditions for Stochastic Dominance are the best known.

Stochastic Dominance (SD) enables uniform ranking of the earnings distributions between females and males ($\ln(w^f)$ and $\ln(w^m)$) over classes of preference functions which include, but are larger than the GE family described above. Let U_1 denote the class of all *increasing* von Neumann-Morgenstern type social welfare functions u that are increasing in wages (i.e. $u' \geq 0$), and U_2 the class of welfare functions in U_1 such that $u'' \leq 0$ (i.e. concave). Concavity implies an aversion to higher dispersion (or inequality, or risk) of wages across individuals:

First Order Dominance:

Male Earnings ($\ln(w^m)$) First Order Stochastically Dominates Female Earnings ($\ln(w^f)$) (denoted $\ln(w^m)$ FSD $\ln(w^f)$) *if and only if*

1. $EF(ln(w^m)) \geq EF(ln(w^F))$ for all $EF \in U_1$ with strict inequality for some EF;
2. Or, $F_2(y) \leq F_1(y)$ for all y with strict inequality for some y.

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

Second Order Dominance:

Male Earnings ($\ln(w^m)$) Second Order Stochastically Dominates Female Earnings ($\ln(w^f)$) (denoted $\ln(w^m)$ SSD $\ln(w^f)$) *if and only if*

1. $EF(ln(w^m)) \geq EF(ln(w^F))$ for all $EF \in U_2$ with strict inequality for some EF;
2. Or, $\int_{-\infty}^{y} F_2(t)dt \leq \int_{-\infty}^{y} F_1(t)dt$ for all y with strict inequality for some y.

Note that $\ln(w^m)$ FSD $\ln(w^f)$ implies that the mean male wage is greater than the female mean wage. Similarly, if ($\ln(w^m)$ SSD $\ln(w^f)$), then the earnings distribution of males is "better" than that of females for all those with *any* increasing and concave evaluation functions in the class U_2 (with strict inequality holding for some utility function(s) in the class). Note that FSD implies SSD. One advantage of this approach is that SD rankings are robust to the wage distributions and/or weights assigned to subgroups within the population. Higher order SD rankings are based on narrower classes of preferences. For instance, Third Order dominance is associated with preference functions with increasing aversion to inequality which place greater weight on equalizing transfers at the lower tails of the earnings distribution. If distributions cross, especially at lower wages, the need for suitable EF functions is indicated which will reflect narrower types of preferences.

In this paper, we employ stochastic dominance tests based on a generalized Kolmogorov-Smirnov test discussed in Linton et al. (2005) and Maasoumi and Heshmati (2000). The Kolmogorov-Smirnov test statistics for FSD and SSD are based on the following functionals:

$$d = \sqrt{\frac{N_1 N_2}{N_1 + N_2}} \min \sup [F_1(y) - F_2(y)] \quad (7)$$

$$s = \sqrt{\frac{N_1 N_2}{N_1 + N_2}} \min \sup \int_{-\infty}^{y} [F_1(t) - F_2(t)]dt \quad (8)$$

The test statistics are based on the sample counterparts of d, and s employing emirical CDFs given by $\widehat{F_1(y)} = \frac{1}{N_1}\sum_{i=1}^{N_1} I(\ln(w_i^f) \leq y)$, where $I(\cdot)$ is an indicator function, and N_1 is the sample size; $\widehat{F_2(y)}$ is similarly defined. The underlying distributions of the test statistics are generally unknown and depend on the data. We use bootstrap implementation of the tests for iid samples. This approach estimates the probability of the statistics 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. We can infer SSD based on s and $Pr[s \leq 0]$ in a similar fashion.[7]

[7]There are equivalent SD conditions expressed in terms of the quantiles. Aaberge et al. (2013) refer to these

4. Data

We examine data for the period 1976-2013, March Current Population Survey (CPS) (available at `http://cps.ipums.org`, King et al., 2010). The March CPS is a large nationally representative household data that contain detailed information on labor market outcomes such as earnings and other characteristics needed for our modeling purposes. It thus has been widely used in the literature to study the gender gap (e.g., Blau and Beller 1988; Mulligan and Rubinstein 2008; and Waldfogel and Mayer 2000). We begin at 1976 since it was the first year that information on weeks worked and hours worked is available in the March CPS. The CPS relies on a complex stratified sampling scheme, hence it is essential to the appropriate weighting variables in our analysis. Following the recommendation of the CPS (King et al., 2010), we use person-level weight (WTSUPP) variable throughout our analysis (which "should be used in analyses of individual-level CPS supplement data").[8]

Our baseline analysis follows the literature closely in defining our variable of main interest and imposing sample restrictions. Following the literature (e.g. Blau and Kahn, 1997), we use the 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. The wages are adjusted for inflation based on the 1999 CPI adjustment factors (available at `https://cps.ipums.org/cps/cpi99.shtml`). Also 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). There are some respondents who do not answer the earnings questions in the March CPS supplement, and the Census Bureau impute wages for these individuals. The literature has shown that *inclusion* of these imputed wages in wage studies is "problematic" (Hirsch and Schumacher 2004; Bollinger and Hirsch 2006). Following the literature (e.g., Mulligan and Rubinstein 2008; Lemieux 2006), we 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).

We examine individuals aged between 18 and 64 who work only for wages and salary and do not live in group quarters. To ensure that our sample includes only those workers with stronger attachment to the labor market, we include only those who worked for more than 20 weeks (inclusive) in the previous year. Moreover, we focus on full-time workers who worked more than 35 hours per week in the previous year, as in the literature (e.g., Mulligan and Rubinstein 2008). Information about sample sizes for each is provided in the supplemental material.

We also repeat all our analysis using NBER Outgoing Rotation Group data to assess the

as "inverse distribution" rankings! Maasoumi (2001) surveys the related tests and techniques. The distribution theory for this kind of testing has been known in statistics and econometrics since before the ones favored in this work. But the asymptotic inferences are known to be challenging and inaccurate due to the dependence on underlying probability functions.

[8]See IPUMS-CPS website for more details.

robustness of our results (See footnote 8), as suggested by a referee.

5. Baseline Results

In this section we present some baseline results for several measures of the gender gap as well as stochastic dominance tests. Stochastic dominance results also highlight some conclusions that cannot be reached without utilizing the corresponding tests.

5.1. Trend of the Gender Gap 1976 - 2013

Panel A of Table (1) reports a number of popular measures of the gender gap. Column (1) displays our metric entropy measure of the gender gap. Recall that, S_ρ is normalized, taking values in $[0, 1]$, and to facilitate the presentation, they are multiplied by 100. The standard errors based on 299 resamples are reported in the supplemental material. Columns (2) and (7) in Table (1) display the gender gap measured as difference of log earnings at the selected percentiles of the log earnings distribution between men and women (mean, 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th}) that are commonly used in the literature.

5.1.1. Discrepancies among the conventional measures

We first note that all conventional measures imply that there exist substantial earnings differentials between men and women. The differences of the log-earnings at the selected percentiles of the earnings distribution between men and women are consistently positive, suggesting that men earn more than women. However, the implied *size* of the gender earnings differentials in the economy vary with the conventional measures. For example, in 1976, the average gender gap at the 10^{th} percentile is about 31 percentage points, while it is about 50 percentage points at the 90^{th} percentile. The difference of the gender gap implied by these two conventional measures is as large as 19 percentage points. The differences at other parts of the earnings distribution indicate the gender gap is between 43 and 46 percentage points. Even though consistently suggesting the existence of the gender gap, none of the conventional measures at a specific part of the distribution seems to represent the gender gap in the rest of the distribution. As discussed above, reliance on the gender gap measured at a particular quantile is not consistent with the weighting schemes implied by any reasonable welfare functions. What is left to be determined, is how to aggregate the different levels of the "gap" at different quantiles. That is, what welfare function weights to use. The "mean" gap uses *equal weights* at all earnings levels and is also sensitive to outliers.

There is a further difficulty with assessment of the conventional measures when one examines the cyclical behavior in the earnings of women relative to men. Looking at the trend from 1976 to 2013, we see a decrease in the difference between the earnings distributions of men and women over the past four decades, regardless of the measure employed. The decrease is not monotonic over time, and the timing of temporal deviations from the long-run trend dramatically varies

Table 1: MEASURES OF THE GENDER GAP

Year	$S_\rho \times 100$ (1)	Mean (2)	10th (3)	25th (4)	50th (5)	75th (6)	90th (7)
			Panel A: Actual Estimates				
1976	10.566	0.432	0.311	0.427	0.461	0.461	0.486
1977	10.110	0.419	0.297	0.397	0.470	0.465	0.470
1978	10.247	0.425	0.272	0.392	0.465	0.472	0.465
1979	10.127	0.421	0.262	0.386	0.468	0.487	0.474
1980	9.857	0.408	0.252	0.382	0.446	0.475	0.448
1981	9.221	0.395	0.263	0.338	0.447	0.477	0.440
1982	8.823	0.394	0.254	0.345	0.448	0.467	0.443
1983	7.645	0.376	0.240	0.323	0.427	0.439	0.451
1984	6.747	0.356	0.247	0.298	0.412	0.427	0.405
1985	6.258	0.343	0.208	0.288	0.357	0.401	0.414
1986	5.607	0.327	0.189	0.297	0.366	0.387	0.393
1987	5.023	0.322	0.214	0.288	0.376	0.365	0.378
1988	4.526	0.305	0.222	0.264	0.329	0.353	0.354
1989	4.215	0.298	0.210	0.244	0.323	0.324	0.352
1990	3.645	0.283	0.183	0.265	0.303	0.336	0.324
1991	3.109	0.254	0.143	0.208	0.273	0.307	0.330
1992	2.784	0.241	0.107	0.211	0.251	0.285	0.299
1993	2.518	0.228	0.145	0.194	0.262	0.274	0.285
1994	2.153	0.215	0.142	0.175	0.235	0.262	0.281
1995	2.048	0.221	0.143	0.197	0.239	0.283	0.276
1996	2.165	0.231	0.148	0.186	0.240	0.248	0.264
1997	2.071	0.227	0.158	0.210	0.241	0.262	0.259
1998	2.082	0.228	0.131	0.214	0.261	0.250	0.274
1999	2.236	0.235	0.154	0.192	0.231	0.251	0.262
2000	2.059	0.232	0.182	0.202	0.247	0.285	0.284
2001	1.913	0.226	0.143	0.192	0.227	0.261	0.300
2002	1.781	0.216	0.117	0.185	0.205	0.236	0.278
2003	1.684	0.208	0.125	0.147	0.203	0.238	0.265
2004	1.353	0.188	0.130	0.153	0.164	0.230	0.288
2005	1.393	0.190	0.105	0.182	0.180	0.219	0.255
2006	1.321	0.191	0.113	0.145	0.182	0.204	0.243
2007	1.102	0.180	0.125	0.153	0.183	0.206	0.227
2008	1.149	0.172	0.118	0.103	0.192	0.214	0.229
2009	1.275	0.186	0.136	0.148	0.163	0.212	0.270
2010	1.275	0.187	0.131	0.134	0.173	0.220	0.254
2011	1.118	0.175	0.111	0.140	0.173	0.211	0.236
2012	1.129	0.178	0.111	0.139	0.170	0.215	0.266
2013	0.974	0.167	0.108	0.113	0.172	0.198	0.223
			Panel B: Average annual percentage changes by Measures				
Pre-1994	-0.080	-0.037	-0.035	-0.043	-0.031	-0.029	-0.030
Post-1994	-0.043	-0.015	-0.007	-0.014	-0.018	-0.014	-0.009
Whole Period	-0.060	-0.025	-0.020	-0.027	-0.024	-0.021	-0.019

[1] Data Source: IPUMS CPS (`http://cps.ipums.org/cps/`). Column (1) reports the entropy gender gap (×100). Columns (2)- (6) report conventional measures.

across different measures used. To ease the presentation, we report the patterns of changes in different measures in Table (A1). The cells with "I" highlighted in green are the years when the measure increased, while the cells with "D" highlighted in light grey are the years when it decreased. As we can clearly see, the conventional measures of the gender gap generally do not move in the same direction together, except in few years (1980, 1989, 1994, and 2002). For example, in 1977, the gender gap at the median and 70^{th} percentile increased , while the gender gap at other selected parts decreased. It would appear that conventional measures fail to be representative of the earnings distribution.

5.1.2. The results using the entropy measure

We first note that S_ρ is statistically significantly different from zero in all cases. S_ρ, provides a summary picture of the trend. Our measure does not employ a fixed weighting scheme, the weights vary over time, which is as expected because of Equations (2)-(4). For example, the entropy gender gap decreased in 1977, consistent with the decrease at all parts but the median and 70^{th} percentile; the entropy gap increased in 1999, consistent with the increase only at the 10th and 75th percentiles. The entropy measure is able to account for *not only* increasing earnings *but also* the greater dispersion (inequality increasing) accompanying it. More important, the entropy measure is consistent with a much broader class of social welfare (evaluation) functions. In 2009 (the great recession), the gender gap implied by the mean increased, while that by the median decreased! This conflicting result could lead to different conclusions about the cyclicality of the gap. In this particular instance, the entropy measure suggests that the gender gap indeed increased statistically (possibly due to worsened economic conditions), in agreement with the mean. One interesting finding worth noting is that the gender gap at the 10^{th} percentile fluctuates more around the trend over time, compared to the other parts of the distribution; the gender gap at the 90^{th} percentile, although fluctuating sometimes, exhibits consistently a declining trend.

5.1.3. The Long-Run Trend Implied by Various Measures

The magnitudes of S_ρ and other measures reported in Table (1) are not directly comparable. To facilitate comparisons of the time patterns we normalize these measures by setting the value of all measures in 1976 to 100, and generate normalized values based on the original growth rates. These normalized values are shown in Figure (1), and the confidence intervals for S_ρ are shaded areas in red.

The results using the conventional measures are broadly consistent with the literature (e.g., Blau and Kahn, 2006). In particular, the gender gap did fall rapidly before the early 1990s, continued in a general downward trend but with a much slower rate until the most recession, and remained relatively stagnant (for most measures) and showed somewhat modest decline (at best) afterward. Moreover, there is larger gender convergence in the lower tail than in the

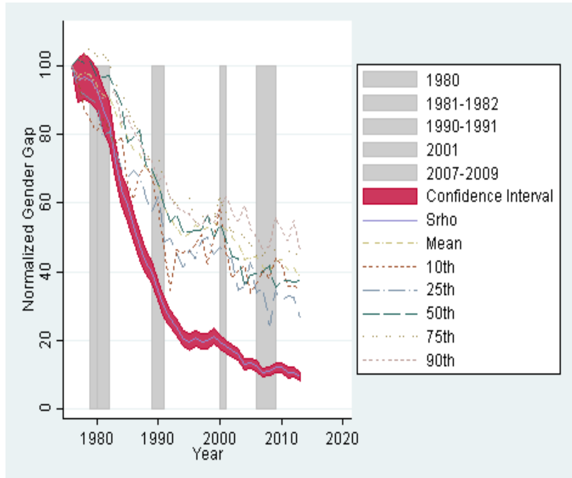


Figure 1: The Trend of Gender Wage Gap (Shaded Areas correspond to the recession periods announced by NBER)

upper tail. In other words, low-wage female workers are catching up more quickly with their male counterparts than high-wage women do.

The pattern implied by our entropy measure is qualitatively similar to that by the conventional measures, in that it also indicates a continuing trend of decline in the gender gap over time. However, the rate of decline implied by our measure is much larger than that by the conventional measures: the gender gap dropped *precipitously* before 1990s, but the convergence drastically slowed down since 1990s. All traditional measures appear to severely *underestimate* the decline in the gender gap over time. Panel B of Table (1) quantifies the differences. In particular, our S_ρ implies the gender gap narrowed at average annual percentage changes of about 6% during the whole period 1976-2013, while the average annual percentage changes implied by the gender difference using conventional measures vary between 2% to 2.7% during the same period. Intuitively, when the gap decreases at many quantiles, a complex weighted measure of them, such as S_ρ, can decrease even more or less notably.

Examining the gender gap separately for different sub-periods, we find that the average annual percentage changes were 8% before 1994, and about 4% afterwards. Although consistent with the literature, this result is a bit surprising. 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 have expected that both welfare reform and the technological progress might accelerate the speed of the convergence. However, our results indicate that this is not necessarily the case, and even if these events helped in some metrics, the effects were offset by some other movements. As we will see below, the *actual* situation is even more prominent once we take into account the women who do not work full-time.

Another interesting phenomenon is observed when we plot the time evolution of these measures against the dates of past recessions as announced by the National Bureau of Economic Research (shaded areas) in Figure (1). Conventional measures of the gender gap respond to business cycles differently, and for each measure, the directions of the responses also vary across time. By contrast, our measure indicates that the gender gap for society as a whole is relatively insensitive to changes in economic conditions, except in 2001. See Table (A1) as well.

5.1.4. The Time Trend Implied by Conventional Measures: Log vs Level

It is important to note a difference between our broad entropy measure and others in terms of standardization. Our measure is invariant to the logarithmic transformation of the earnings series since, as was stated earlier, it is invariant to all monotonic (linear or non linear) transformations. There is no need to reinterpret according to data transformations. This is not so for all the other metrics in this table, since they will change depending on whether one uses the actual earnings series, or their logarithm, or some other transformation. As suggested by a referee, we also plot the (normalized) conventional measures of the gap using both the levels and logs of wages in the extended working paper version of this paper. While the implied time trend by the gap at $10^{th}, 25^{th}$ and 50^{th} percentiles is relatively similar between levels and logs, the rate of decline implied by the measures using the levels is slightly larger than that implied by using logs; this is particularly true for the mean gender gap. In this context, the discrepancies between using levels and logs are relatively modest. However, it is conceivable that in other contexts when such differences may be large, our entropy measure would be more desirable.

5.2. Stochastic Dominance Results

Here we report observed distributions of earnings for men and women without any controls for age, education, or other relevant factors. SD results are similar across years, and we summarize the results here. In the interest of space, we report only the most recent year in Table (2), as an example. Other SD results are available in the extended working paper and the graphical comparisons of CDFs in the supplemental material. Note that the column labeled *SD* details if the distributions can be ranked in either the first or second degree sense; 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 dominance to a desirable degree of confidence.

The graphical comparisons of the CDFs show that the earnings distribution for men lie predominantly to the right of the one for women, indicating higher level of earnings for men. This observation is consistent with the fact that the differences in selected percentiles of the earnings distributions between men and women are uniformly positive. Moreover, the earnings distributions between men and women are getting closer over time, in line with the results implied by our measure of the gender gap.

Table 2: An Example of Stochastic Dominance Results (Female v.s. Male Wage Distributions)

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

[1] We find first-order dominance in all years with probability 1. The equality in the limits of CDFs at the end points is anticipated by the test procedure. In the interest of space, only the most recent year is reported as an example. Other results are available in the extended working paper version.

Our SD test statistics confirm these graphical observations. All the SD results are similar to the example reported in Table (2): we do observe stochastic dominance throughout the whole period. In particular, the earnings distribution among men is found to empirically dominate, in a first-order sense, the earnings distribution among women in every case! This result is extremely powerful: any individuals with a social welfare function in the class U_1 (merely increasing in earnings) would prefer the male distribution to the female distribution, concluding that men perform better in the labor market and hence enjoy a higher-level of well-being than women. Moreover, all these observed dominance relations are highly statistically significant, with probability 1 across all years. Despite the fact that we have observed the narrowing gap of the two distributions in the past four decades, women do not perform better than men across the whole distribution. Later in this paper we report results with controls for some characteristics.[9]

Dominance orderings indicate that women are generally uniformly less well off than men, although women's earnings distribution may have improved relative to men's. This is generally

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

consistent with the previous literature such as Blau and Kahn (2006). We strongly qualify these findings due to lack of control for covariates and for selection.

6. Accounting for Selection

Not all women are full-time employed or even participate in the labor market. Past studies have documented that the female labor force participation, although has been rising over time, remains low compared to men, and women in other developed countries (Blau and Kahn 2013). This is also true for full-time employment among women. Using our own data, we also document that despite a remarkably rapid increase from the beginning of our sample through 2001, the women's full-time employment remains at roughly 53%. Interestingly, the trend has reversed after 2001, which is associated with the increase in stay-at-home mothers (Cohn et al. 2014). As a result, analysis based only on the full-time female workers may not be informative and could even be misleading for assessing the population gap.

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. Both authors find the 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 wage observations 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)). In other words, missing wages are filled in based on observable characteristics. In addition, Olivetti and Petrongolo (2008) also implicitly assume a fixed selection rule, which in reality may fail to hold, to interpret their results. By contrast, Mulligan and Rubinstein (2008) allow for selection on unobservables and time-varying selection pattern. Their focus is on the selection bias in measuring the *mean* gender gap using the conventional (i.e., parametric) Heckman type of selection model. They similarly find that the selection plays an important role in explaining the observed gender gap, and that the selection pattern indeed varies over time, from negative to positive. Instead of making a parametric selection correction, Blundell et al. (2007) uses economic theory to derive bounds on the gender wage gap. Unlike Mulligan and Rubinstein (2008) and Olivetti and Petrongolo (2008) who focus on either mean or median, Blundell et al. (2007) are also able to derive the bounds for the gender gap across different parts of the *whole* distribution. Their key assumption is the fixed, positive selection assumption that working women's wages first order dominate non-working women's wages. In other words, employed women have higher wage offers than non-employed women. This assumption may be too restrictive and fail to hold in reality. And indeed, evidence of negative selection has been documented in Neal (2004) and Mulligan and Rubinstein (2008).

To the extent that the wages among non-working women are systematically different from full-time workers, our measures of the gender gap and stochastic dominance results may under- or over-estimate the well-being of women relative to men. For example, if the wages among non-working women are systematically lower than those among full-time workers (i.e., positive selection), we under-estimate the gender gap while the stochastic dominance relations may remain unchanged and even strengthened. But if the wages among non-working women are higher than full-time workers (i.e., negative selection), we over-state the severity of women's labor market performance relative to men.

While econometric tools are well developed and relatively mature for addressing selection in conditional *mean* models (such as Heckman type of models, semi-parametric or not), such tools are sparsely developed for our purpose — namely addressing selection at the distributional level and recovering the *unconditional* wage distributions. Without explicitly addressing the selection, the best we can do is identify the distributions for the selected sample only. Our measures and analysis of the gender gap in earlier sections can be considered as the gender gap for full-time workers only.

6.1. Econometric Methods to Address Selection

To address the selection issue at the distributional level and allow for a time-varying selection pattern, we adopt a two-step procedure that recovers the marginal distributions of the wages based on conditional quantiles. Machado and Mata (2005) is among the first empirical studies that utilize such a procedure. The asymptotic statistical theory has recently been developed in Chernozhukov et al. (2013). Our procedure differs from Machado and Mata (2005) and some subsequent analysis in that we take into account selection when estimating the conditional quantiles. While the econometric tools are well developed and relatively mature for addressing the selection issue for conditional mean models in the tradition of Heckman, there are a few important developments for inference on other parts of the conditional distributions. There are (to the best of our knowledge) three methods proposed in the literature to *point* identify the conditional quantile models in the presence of selection. As will be reviewed in more detail later, the approach proposed in Arellano and Bonhomme (2012) has many advantages and is indeed the one adopted in our paper. Their approach is a semi-parametric approach that ingeniously models the joint distribution of the true (or latent) quantile of the wage distribution and participation decision, while leaving their marginal distributions unspecified. Their approach allows the parameter capturing the selection to be estimated and, unlike Blundell et al. (2007), it does not impose any restrictions on the selection. We are thus able to assess the magnitude of the selection as well as the change of the selection pattern over time. This result could also help explain the pattern in the *observed* gender gap. To account for selection we proceed in two steps. We first estimate conditional quantile functions in the presence of selection, and then recover the unconditional distributions based on these functions. This procedure differs from the literature (e.g., Machado and Mata 2005) in that the first stage takes into account the selection issue.

6.1.1. *Preliminaries: Recovering Unconditional Distributions from Conditional Quantile Functions*

To summarize the relationship between *conditional* quantiles and *unconditional* distributions, let $F_{y|x_i} \equiv Pr[y_i \leq y|x_i]$ be the conditional CDF of the wages given $x = x_i$, and $Q_\tau(y|x_i)$ the corresponding τ^{th} conditional quantile. Note that $Q_\tau(y|x_i) = F^{-1}_{y|x_i}(\tau)$, the inverse of the conditional CDF. First note that the marginal distribution and the conditional distribution are related by[10]

$$F_y(y) = \mathbb{E}[I[y_i \leq y]] = \mathbb{E}[\mathbb{E}[I[y_i \leq y]|x_i]] = \mathbb{E}[F_{y|x_i}] \tag{9}$$

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

$$F_{y|x_i} = \int_0^1 I[F^{-1}_{y|x_i}(\tau) \leq y]d\tau = \int_0^1 I[Q_\tau(y|x_i) \leq y]d\tau \tag{10}$$

The intuition behind this is that the conditional quantiles over the support $(0, 1)$ characterizes the *whole* data process. Together, Equations (9) and (10) imply that

$$F_y(y) = \mathbb{E}[F_{y|x_i}] = \mathbb{E}[\int_0^1 I[Q_\tau(y|x_i) \leq y]d\tau] \tag{11}$$

As suggested in Machado and Mata (2005), the unconditional CDF could thus be estimated by

$$\widehat{F}_y(y) = \frac{1}{N}\sum_{i=1}^{N}\int_0^1 I[Q_\tau(y|x_i) \leq y]d\tau] \tag{12}$$

And the corresponding unconditional quantile can be obtained by inverting the CDF:

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

Using these estimates, we can easily calculate the gender gap proposed above, as well as test for stochastic dominance.

6.1.2. *Conditional Quantile Selection Models*

As we can see in (12), to obtain the quantities of interest we need the conditional quantile function $Q_\tau(y|x_i)$. Parametric estimation of the conditional quantile functions has been developed since Koenker and Bassett (1978), and nonparametric extensions have recently been pro-

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

posed (e.g., Li and Racine 2008). Note that in the absence of selection, a re-weighting approach can also be used to recover the 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 because it is essentially a nonparametric approach. As noted in Newey (2007) and Huber (2014), without further assumptions, we cannot identify the distributions for other groups than the selected sample, mainly because we do not observe the wages for people who do not work. On the other hand, the quantile function approach, by adding more structure and hence information, opens a door to model and address the selection issue and enables identification. We believe this is the only feasible way in our context at this point. Our approach here is a semi-parametric method which strikes a good balance between imposed structure and flexibility.

There are (to the best of our knowledge) only three approaches available to *point* identify the parameters in the conditional quantile functions – identification at infinity, the Buchinsky (1998) approach, and the Arellano and Bonhomme (2012) approach.[11] 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.

Buchinsky (1998) is the first to propose a control function type of approach (similar to Heckman type of solutions) to address the selection issue in the context of quantile function estimations. That paper assumes that wages are a separable and additive function of observable and unobservable characteristics. More important, it also implicitly assumes "independence between the error term and the regressors conditional on the selection probability." (Melly and Huber 2008) This assumption leads to parallel quantile curves, in other words, all quantile functions are identical and equal to the conditional mean function (Melly and Huber 2008; Arellano and Bonhomme 2012). It is then not evident why quantile regression is needed. As noted in Arellano and Bonhomme (2012), it is indeed difficult to specify a data generating process consistent with this approach.

[11] Olivetti and Petrongolo (2008) propose another approach but focusing only on median regressions. More important, 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 people who never work.

By contrast, Arellano and Bonhomme (2012) propose an ingenious way to address the selection issue but overcome some of the shortcomings of the other two approaches discussed above. In the presence of selection, their approach entails shifting the percentiles as a function of the amount of selection. In the absence of selection, it is just standard quantile regression.

To see how it works, consider the wage function in the form of the conventional quantile model (see, e.g., Chernozhukov and Hansen 2008)

$$\ln(w) = g(x, u) \quad u|x \sim Uniform(0, 1) \tag{13}$$

where $\tau \mapsto g(x_i, \tau)$ is strictly increasing and continuous in τ. Note that this is a non-separable function of observable characteristics, x, and unobservable u. The unobservable disturbance u is normalized and typically interpreted as ability (Doksum 1974; Chernozhukov and Hansen 2008). Equation (13) implies that $g(x, \tau)$ is the τ^{th} quantile of $\ln(w)$ conditioned on x, by construction.[12]

An individual's decision to take full time employment is given by

$$\begin{aligned} S &= I(\ln(w) - R(z) - \eta \geq 0) \\ &= I(g(z) - \eta \geq 0) \end{aligned} \tag{14}$$

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

As commonly used in the treatment effects literature, Equation (15) can be re-written as follows

$$S = I(v \leq p(z)) \quad v|x \sim Uniform(0, 1) \tag{15}$$

where $p(z) = \Pr[S = 1|z]$ is the propensity score satisfying the assumption that $p(z) > 0$ with probability one.[13]

In the absence of selection, the τ^{th} quantile of the wage distribution, $g(x, \tau)$, satisfies the

[12] $\Pr[\ln(w) \leq g(x, \tau)|x] = \Pr[g(x, u) \leq g(x, \tau)|x] = \Pr[u \leq \tau|x] = \tau$. The first equality follows from Equation (13). The second follows from the fact that conditional on x, u is uniformly distributed, a normalized unobservable variable.

[13] Note that Equation (15) is a normalization commonly used in the treatment effects literature. This result is due to the fact that $\mathbb{E}[S|z] = \Pr[S = 1|z] = p(z) = \mathbb{E}[S = 1|p(x)] = \Pr[S = 1|p(z)]$. See, e.g., Imbens (2004, p.9) for a proof.

following $\Pr[\ln(w) \leq g(x,\tau)|x] = \tau$ (see footnote 12). We can estimate $g(x,\tau)$ by standard quantile regression. 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 τ in the selected sample. Instead, the observed rank is $G_x(\tau, p(z))$.[14] If we were to know the mapping between the quantile and its observed rank in the selected sample, we can then estimate $g(x,\tau)$ using "rotated quantile regression". This is indeed the idea proposed by Arellano and Bonhomme (2012). The authors show that both exclusion restrictions and functional forms regarding $G(\cdot)$ provide identification. Instead of simply relying on the assumption of functional forms on $G(\cdot)$, we also use exclusion restrictions (typically used in the literature) to help identify $G(\cdot)$ and hence the conditional quantile function (see below for further details). Although we need to specify a copula, the marginal distribution of each argument in it is unspecified and can be nonparametric.

6.1.3. Practical Implementation

In practice, propensity scores are estimated using probit models. To allow for flexible specification of the propensity scores, we include in estimation the polynomial terms of the continuous variables up to third order, as well as the interaction terms between them and other discrete variables, in addition to the IV. These variables enter the probit models additively and hence our model takes the form of a standard sample selection model. Such a linear index model makes "identification and estimation simple and transparent." (Bonhomme et al. 2014)

Following the standard practice in the literature, we assume a linear conditional quantile function $g(x,u) = x'\beta(u)$. We want to emphasize that despite the linearity for a given quantile, this model is not as restrictive as it may seem. It is still *nonlinear* in nature because this specification allows x to have differential impacts on the wage distributions.[15] And it is a non-separable function of x and u, allowing for interaction between the observable and unobservable characteristics, and is thus preferred to the additive structure that is often assumed in the conditional mean models. Moreover, Angrist et al. (2006) show that linear quantile regression provides a weighted least squares approximation to an unknown and potentially nonlinear conditional quantile regression. Below we provide some graphic evidence of the performance of such linear yet non-separable models.

[14] To fix ideas, consider a simple (yet unrealistic) numerical example. Suppose that the conditional distribution of log wages with 9 possible values being {1, 2, 3, 4, 5, 6, 7, 8, 9}, while we only observe the selected sample with the wages greater than or equal to 5, i.e., {5, 6, 7, 8, 9}. In the absence of selection, the median ($\tau = .5$) is 5. But its observed rank is instead .10 in the selected sample.

[15] 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".

Following Arellano and Bonhomme (2012), we use a Frank copula in our analysis, which "has been widely used in empirical applications" (Meester and MacKay 1994; Trivedi and Zimmer 2005). This copula is indexed by a parameter, ρ, that captures the dependence between the marginal distributions of the error terms, i.e., $G_x(\tau, p(z)) \equiv G_x(\tau, p(z); \rho)$. As noted in Trivedi and Zimmer (2005), the Frank copula is popular "for many reasons". Among them, it is "comprehensive" because it permits a wide range of potential dependencies. Also, unlike some other copulas, it permits negative dependence between the marginal distributions. As a result, the dependence parameter ρ has a convenient interpretation as it can indicate the sign of selection. A *negative* ρ indicates *positive* selection into employment, while *positive* ρ implies *negative* selection. Even though the copula between the error terms in the selection and outcome equations is parametric, the marginal distributions are not and may be nonparametric (Huber and Melly 2012).

As in Arellano and Bonhomme (2012) we employ a three-step procedure: 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 "rotated quantile regression". For our purpose of recovering the unconditional distribution, we estimate β_τ for $\tau = 0.02, 0.03, \ldots, 0.97, 0.98$.[16]

6.2. Data and Exclusion Restrictions

In our analysis, 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. 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.

Similarly, our choice of exclusion restrictions for the female labor supply equation is a popular one in the literature. There are two commonly used variables in the literature: husband's

[16]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.

income and the number of young children (e.g., Mulligan and Rubinstein 2008; Machado 2012; Buchinsky 2001; Chang 2011; Martins 2001). For example, Mulligan and Rubinstein (2008) use the number of children younger than six, interacted with marital status as variables determining employment, but excluded from the wage equation. 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. Indeed, Huber and Mellace (2011) test the validity of these two exclusion restrictions used in the sample selection models. 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". Considering this tradition and empirical evidence, we use the number of children under age 5 as an exclusion restriction.

6.3. Results

It would be useful to have a sense of how well the quantile selection models perform under these assumptions. One way to do so is to compare the wage distribution of the full-time workers recovered based on quantile selection models, with the wage distribution that is actually observed in our sample. Figure (2) displays this comparison at a few quantiles, $\tau = .10, .25, .50, .75, .90$.

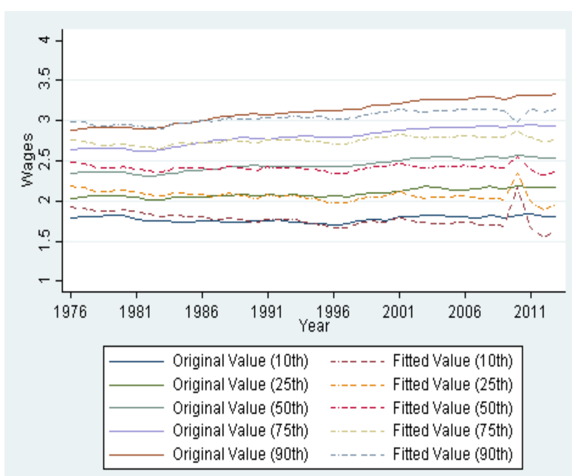


Figure 2: The Goodness of fit for Quantile Selection Models: the Quantiles of Full-time Workers with and without correction for selection

The quantile models perform relatively well here. While there are inevitable specification errors in these models, in most years, specification errors are within a very small neighborhood. In many years, the imputed quantiles based on quantile selection models are even identical or close to identical to the observed ones. While this certainly does not constitute a formal test of all the assumptions, such results do provide some confidence in the methods used here and the results that follow. As we will show below, both of our "derived" mean and median gap measures are close to those during the periods obtained by Mulligan and Rubinstein (2008) and Olivetti and Petrongolo (2008) using completely different approaches. [17]

6.4. The Gender Gap After Controlling for Selection

6.4.1. Selection and The Magnitude of the True *Gender Gap*

The results are presented in Table (3). After correction for selection, we continue to find positive differentials at the select earnings percentiles between men and women. This finding confirms the existence of the gender gap and again suggests that men typically earn more and fare better than women.

Regardless of the measure used, the estimated size of the gender gap after the selection issue is controlled for is in stark contrast to the results for the working sample. (Below, we denote the gender gap without addressing the selection as either the *raw*, or the *observed* gender gap; we denote the gender gap addressing the selection as the *true* or *population* gap.)

Two findings are notable. First, the differences in the results before and after addressing the selection vary in terms of both magnitude and direction over time. In early years, for example, in 1976 and 1977, the population gender gap is smaller than the observed gap, regardless of the gap measure. The difference with and without selection can be as large as 12 percentage points at the 10^{th} percentile, implying that the true gender gap is overestimated by roughly 38 percent. On the other hand, in many later years, the true gender gap is much greater than the observed one.

Second, the differences of the gender gap before and after addressing the selection also vary in terms of magnitude and direction across the distribution. For example, in 1983, the population gender gap is underestimated by the working sample at all selected percentiles except the 10^{th} percentile where it is over-estimated. In 1984, the population gap is overestimated everywhere, except at the 90^{th} percentile. Altogether, these results suggest that the selection issue is indeed significant, and that failure to take it into account can severely bias the estimates of the gender gap. More important, these results also suggest that the selection pattern has been changing over time as well as across the distribution.

These conclusions are indeed corroborated by the estimated dependence parameter, ρ, in

[17]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, and could thus be numerically less stable. We should therefore be cautious when interpreting the results for this particular year. In what follows, we also exclude the year 2010 to better discern the patterns when needed.

Table 3: Measures of The Gender Gap (Corrected for Selection)

Year	$S_\rho \times 100$ (1)	Mean (2)	10th (3)	25th (4)	50th (5)	75th (6)	90th (7)
1976	8.179	0.321	0.194	0.303	0.344	0.363	0.393
1977	9.216	0.338	0.203	0.299	0.383	0.387	0.417
1978	11.175	0.391	0.219	0.364	0.432	0.451	0.460
1979	11.115	0.393	0.219	0.352	0.445	0.464	0.465
1980	10.096	0.365	0.197	0.325	0.406	0.442	0.437
1981	9.633	0.354	0.188	0.301	0.401	0.434	0.433
1982	10.041	0.368	0.198	0.307	0.423	0.454	0.450
1983	10.373	0.397	0.238	0.331	0.448	0.471	0.501
1984	7.896	0.339	0.193	0.285	0.382	0.413	0.425
1985	8.356	0.356	0.193	0.294	0.378	0.433	0.443
1986	7.624	0.347	0.183	0.307	0.385	0.417	0.433
1987	8.577	0.398	0.263	0.351	0.448	0.455	0.464
1988	6.905	0.340	0.215	0.284	0.365	0.399	0.410
1989	7.413	0.361	0.228	0.315	0.398	0.411	0.433
1990	7.323	0.383	0.266	0.350	0.415	0.445	0.442
1991	5.423	0.309	0.179	0.262	0.331	0.359	0.396
1992	5.478	0.316	0.165	0.279	0.340	0.374	0.384
1993	5.191	0.308	0.183	0.274	0.339	0.365	0.383
1994	5.025	0.317	0.208	0.273	0.338	0.382	0.386
1995	5.043	0.333	0.236	0.299	0.353	0.386	0.397
1996	5.913	0.374	0.256	0.348	0.384	0.388	0.406
1997	5.870	0.372	0.268	0.350	0.378	0.397	0.413
1998	5.243	0.346	0.218	0.332	0.376	0.365	0.402
1999	5.205	0.343	0.245	0.295	0.330	0.358	0.406
2000	5.131	0.347	0.252	0.305	0.363	0.392	0.403
2001	4.307	0.310	0.202	0.258	0.296	0.355	0.404
2002	5.365	0.366	0.237	0.335	0.357	0.382	0.444
2003	5.949	0.394	0.293	0.362	0.396	0.407	0.463
2004	5.351	0.375	0.287	0.345	0.364	0.393	0.469
2005	5.325	0.372	0.267	0.350	0.366	0.396	0.441
2006	4.624	0.344	0.254	0.297	0.321	0.363	0.415
2007	4.796	0.362	0.280	0.334	0.361	0.383	0.424
2008	5.165	0.374	0.319	0.326	0.386	0.401	0.433
2009	5.579	0.393	0.327	0.362	0.376	0.408	0.452
2010	19.034	0.276	-0.044	0.037	0.280	0.403	0.590
2011	6.845	0.451	0.425	0.443	0.442	0.451	0.477
2012	8.409	0.517	0.487	0.533	0.501	0.494	0.541
2013	6.572	0.452	0.421	0.445	0.445	0.451	0.485

[1] Data Source: IPUMS CPS (`http://cps.ipums.org/cps/`). Column (1) reports the entropy gender gap ($\times 100$). Columns (2)- (6) report conventional measures.

Table (4). This parameter also varies in terms of magnitudes and directions over time. The transition over time does not occur immediately. It was mostly positive in the early period (up to1991), is close to zero in 1992 and increasingly negative from then on. Recall that the positive dependence indicates negative selection, while negative dependence suggests positive selection. Therefore, the gradually declining trend of the dependence parameter is consistent with the observed change in the selection pattern, from negative to positive. 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).[18] Neal (2004) similarly emphasizes that the selection pattern can be either positive or negative. Below we discuss two reasons for the observed transition in the selection pattern.

Table 4: Estimated ρ and Implied Pattern of Selection into Employment

Year	ρ	Implied Selection	Year	ρ	Implied Selection	Year	ρ	Implied Selection	Year	ρ	Implied Selection
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			

[1] **N** denotes negative selection, while **P** denotes positive selection.

Another result is worth emphasizing. Although using completely different approaches, our estimates of both the mean and median gaps 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 column (2) of Table (1). 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. These estimates are indeed similar to our results presented in column (2) of Table (3). 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

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

imputation method based on wage observations from adjacent waves range from .339 to .363 (in their Table 2), and the results using the imputation method based on observables from a probabilistic model range from .359 and .371. These estimates are similar to our results during the same period (column (5) of Table (3)), ranging from .330 to .384. The fact that these results are also in line with the literature provides some confidence in our results for the rest of the distribution.

6.4.2. The Trend of The Population Gender Gap

The results also suggest that the "evolution" of the gap is influenced by selection. For example, 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. To visualize our results, we again construct the normalized values of different measures, and these results (without year 2010) are presented in Figure (3). Analogous to Table (A1), Table (A2) summarizes the the patterns of changes in different measures. Similar to Mulligan and Rubinstein (2008), we find that the mean gender gap narrowly fluctuated around the starting level during the 1970s through 1990s. While the timing of the fluctuations is different for the median gender gap, it too oscillates around its own starting level during this period. However, this is not the case for some other parts of the distribution. For example, we find that the gender gap at the 10^{th} percentile began to increase and deviated greatly from its starting level in early 90s. These quantile specific observations are informative but clearly fail to be representative. A representative notion of the overall "gap" must reveal its weighting scheme.

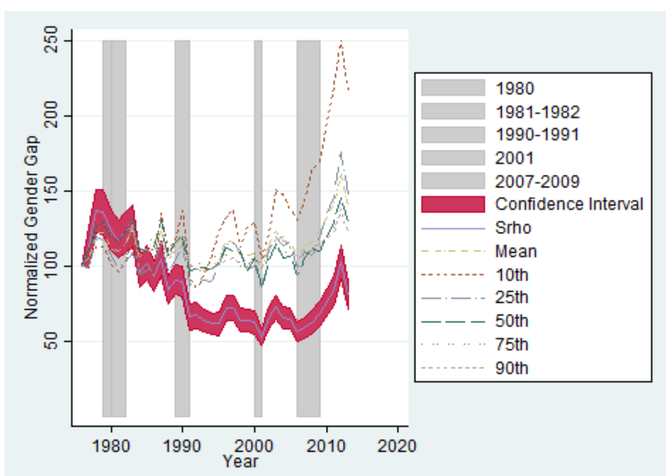


Figure 3: The Trend of Gender Wage Gap (Corrected for Selection into Employment)
Note: Year 2010 is excluded to better reflect the time trend.

The disagreement between different gap measures at different parts of the distribution highlights, again, the challenge of selecting narrow distribution measures like the mean or any single quantile. In the early years, there is no dominance order, making the choice of a summary measure both necessary and sensitive. We prefer a measure like S_ρ which accounts for all the moments of the distribution as well as inequality/dispersion. In particular, our measure indicates that the population gender gap in the U.S. may have decreased between 1970s and early 1990s, although at a slower rate than suggested by the employed sample (as evident by the comparison of Figures (fig.overall) and (3)). The gap was stable until the most recent recession.

The pattern since 2007, the start of the recent recession, is worth noting. The population gap soared from the beginning of the recession, while the working sample gap remained relatively stable. Such a difference indicates that the recent recession may have hurt low-skilled women's labor market prospects and "forced" them out of full-time employment. Women who stay in full-time employment tend to the higher-wage earners (i.e., a positive selection),. The working sample gender gap underestimates the true gap.

The results presented here tend to contradict some of the earlier findings in the existing literature. The entropy measure suggests that, although there was some narrowing of the gap in early 1990s, such progress was much smaller than previously found in the literature that fails to take into account women's selection into employment. Also, the commonly found "slower" convergence (in the wage distributions between men and women) during the 1990s is even slower with the entropy measure. This is because it accounts for "inequality" and dispersion. This has implications for the analysis of policies aimed at improving women's wellbeing. Interestingly, women's worsened situation during the recent recession is missed by analyses that only observe women earning higher wages in the labor market.

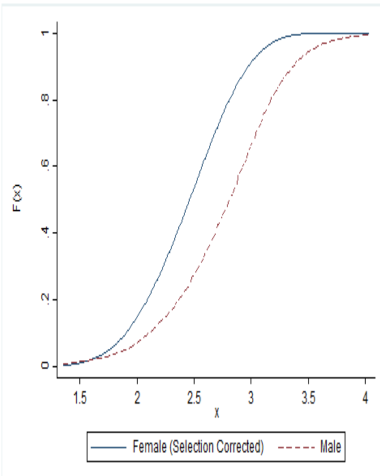
6.5. Stochastic Dominance Results

Negative selection would imply crossing of distributions at lower wages. Dominance relations are then less likely, a priori, once selection is controlled for. A positive selection, on the other hand, may strengthen the earlier findings of SD relations between the full time employed.

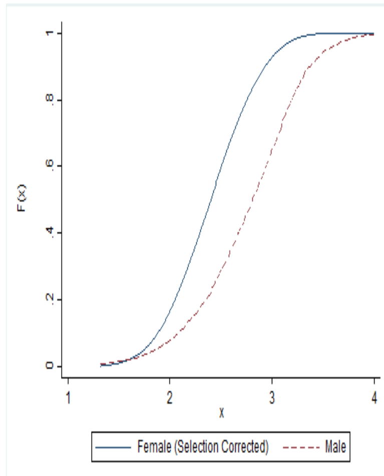
Table (5) summarizes the outcome of statistical SD tests, corrected for selection. Given the relatively large decline in the gender gap during the early years, it is less likely to find statistically significant first-order dominance, or any higher order dominance relations before the early 1980s. There is no statistical dominance relations in 4 cases (1976, 1977, 1980, and 1981) and none in 1979 and 1982 (FSD is at low degree of confidence, respectively 0.89 and .82). The inability to rank order the earnings distributions between men and women in this case is equally informative. It implies that for 1998, say, 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.

In contrast the mere observation of the gap at specific percentiles of the distributions is bewildering. Recall that we find positive differences in wages in favor of men at certain percentiles. We plot the CDF comparisons for select years in Figure (4) and the results are for other years are available in the supplemental material.

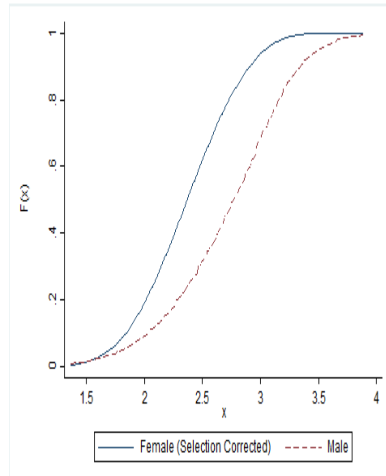
These graphs are illuminating. For all the no dominance cases, there is an early crossing



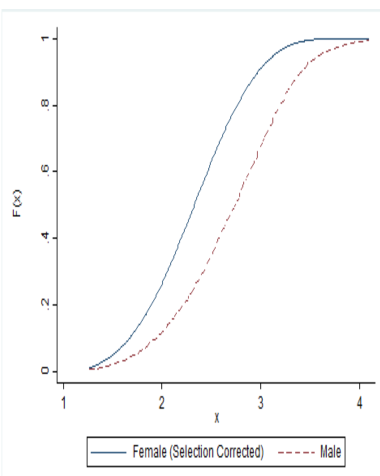
(a) 1976



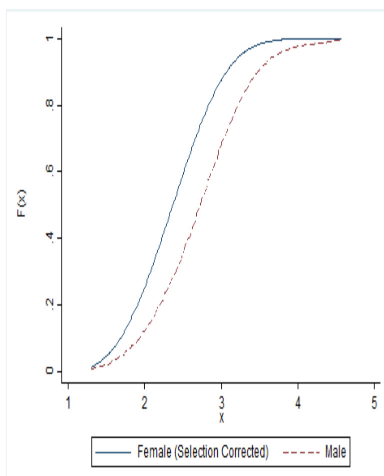
(b) 1980



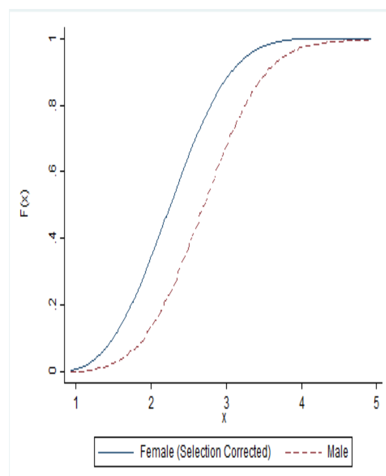
(c) 1981



(d) 1990



(e) 2000



(f) 2013

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

Table 5: 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	N	0.74	0.67	15.65	0.67	1995	F	-0.83	1.00	-0.83	1.00
1977	F	1.09	0.12	35.34	0.12	1996	F	-0.81	1.00	-0.81	1.00
1978	F	-0.78	0.99	-0.78	0.99	1997	F	-0.81	1.00	-0.81	1.00
1979	F	-0.84	0.89	-0.92	0.89	1998	F	-0.84	1.00	-0.84	1.00
1980	N	0.97	0.37	24.66	0.37	1999	F	-0.80	1.00	-0.94	1.00
1981	N	0.98	0.49	12.98	0.49	2000	F	-0.88	1.00	-0.88	1.00
1982	F	-0.81	0.82	-0.85	0.82	2001	F	-1.08	1.00	-1.08	1.00
1983	F	-0.79	1.00	-0.87	1.00	2002	F	-1.02	1.00	-1.02	1.00
1984	F	-0.78	0.93	-0.89	0.93	2003	F	-1.01	1.00	-1.14	1.00
1985	F	-0.80	0.94	-0.81	0.94	2004	F	-1.07	1.00	-1.07	1.00
1986	F	-0.80	0.90	-0.80	0.90	2005	F	-0.92	1.00	-1.00	1.00
1987	F	-0.79	1.00	-0.79	1.00	2006	F	-1.07	1.00	-1.07	1.00
1988	F	-0.87	1.00	-0.90	1.00	2007	F	-1.03	1.00	-1.03	1.00
1989	F	-0.56	0.97	-0.84	0.97	2008	F	-0.99	1.00	-1.06	1.00
1990	F	-0.91	1.00	-0.91	1.00	2009	F	-0.83	1.00	-1.02	1.00
1991	F	-0.87	1.00	-0.93	1.00	2010	N	5.98	1.00	340.39	1.00
1992	F	-0.87	0.99	-0.99	0.99	2011	F	-0.96	1.00	-0.98	1.00
1993	F	-0.84	1.00	-0.93	1.00	2012	F	-0.96	1.00	-0.98	1.00
1994	F	-0.86	1.00	-0.86	1.00	2013	F	-0.97	1.00	-1.02	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.

of the CDFs, while the CDF of men's wage distribution lies predominantly under that of women's elsewhere. At the extreme lower tail, women perform slightly better than men, while other women fare worse than men. This result is indeed the motivation for why we adopt our entropy measure and the SD approach to study the gender gap. It illustrates and underscores the benefit to considering the *entire* distribution within a decision theoretic framework, and it also highlights what could be missed should we simply look at select parts of the wage distributions. A narrower class of preference functions would order these distributions. These must be "increasingly averse" to inequality at lower or higher ends of the earnings distribution. Indeed one can see how an "upward" aversion to inequality, as described recently in Aaberge et al. (2013) may rank these crossing distributions. The class of functions that may uniformly rank distributions that cross entail narrower and increasingly "non-consensus" interpersonal comparisons of well being, such as embedded in the Gini type of preference functions. In such situations it is more than usually important to be explicit about the properties of any evaluation function employed to characterize the gap and make decision on it.

In the period beyond the late 80s (except for 2010) men's earnings first-order dominate women's in nearly all cases to a high degree of statistical confidence. Nevertheless, the distance between men and women wage distributions (our entropy measure) decreased before the most recent recession.

Has women's earnings distribution improved over time? As noted in Blau (1998), this receives less attention than the trends in the gender differences. SD tests and plots of the CDFs of the female wage distributions between 1976 (the starting year) and more recent years is given in Figure (5); the SD test statistics are reported in the supplemental material. We exemplify by three pairs of comparisons: 1976 vs 2001 (2006, 2013). Year 2013 is the most recent year available, 2006 is the most recent year before the recession, and 2001 is the year when the gender gap is smallest. We observe that women's wage distribution in 1976 actually dominates, in a second-order sense, women's wage distribution in all three (later) years. In other words, women's labor market outcomes have deteriorated over time (should we take into account "dispersion" in the respective distributions). Examining Panel A of Figure (5), we can see that, despite an improvement in the upper, women in the lower tail of the distribution have lost ground over time. This result is in stark contrast to the findings without addressing the selection that women are found to improve over time at every part of the distribution (i.e., there is observed first-order dominance relations between later and early years) (Panel B of Figure 5); this observation is consistent with Blau (1998).

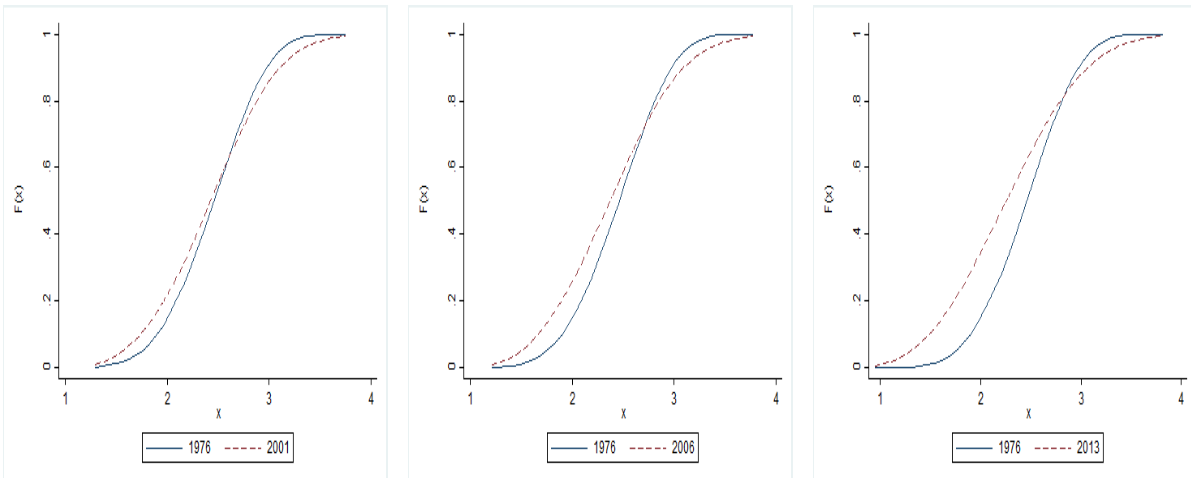
6.6. Why the observed changes in the selection pattern?

Why do we observe a transition from negative to positive selection over time? Here we offer two explanations.

6.6.1. Explanation 1: An Increase in Within-Group Inequality

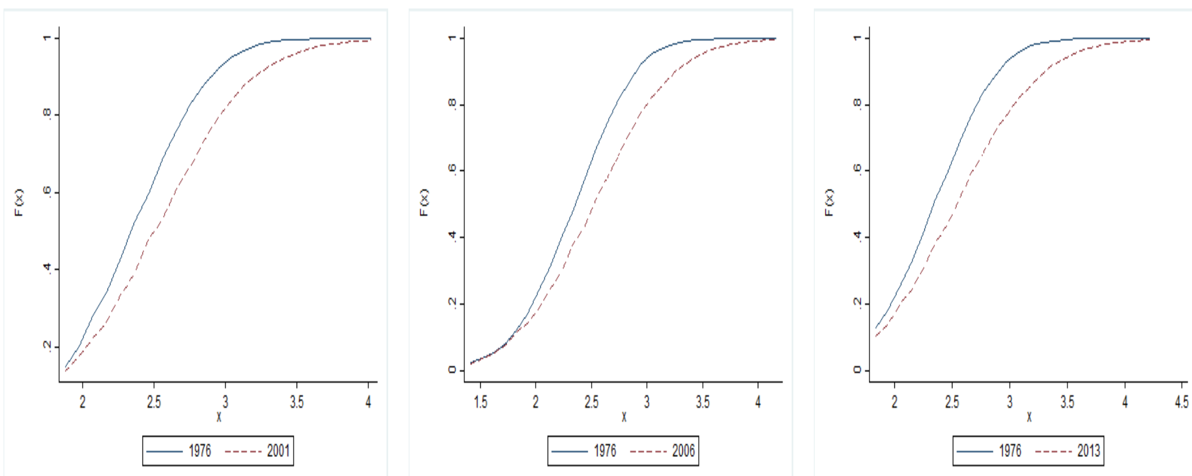
The first explanation is put forth in Mulligan and Rubinstein (2008). Using the traditional Roy model, they argue that increasing wage inequality within gender over time would cause women to invest more in their market productivity and lead abler and hence higher-wage women to participate in labor force. In their paper, the theory is based on the underlying wage inequality, while their explanation is based on the observed wage inequality. These two quantities could be drastically different.

Because our approach recovers the distributions of the potential wage outcomes, this allows us to formally test this explanation. Here we provide the dynamics of the wage inequality measures used in Mulligan and Rubinstein (2008) in Figure (6): the difference of the 90^{th} and 10^{th} percentiles of the log wages. In addition, we also plot the difference of the 50^{th} and 10^{th} percentiles, as well as the difference of the 50^{th} and 10^{th} percentiles. Consistent with Mulligan and Rubinstein (2008), we find that the *observed* within-group inequality for both men and women were flat before 1980 but increased rapidly during the period between early 1980s and



(a) 1976 vs 2006 (b) 1976 vs 2006 (c) 1976 vs 2013

Panel A: Corrected for Selection

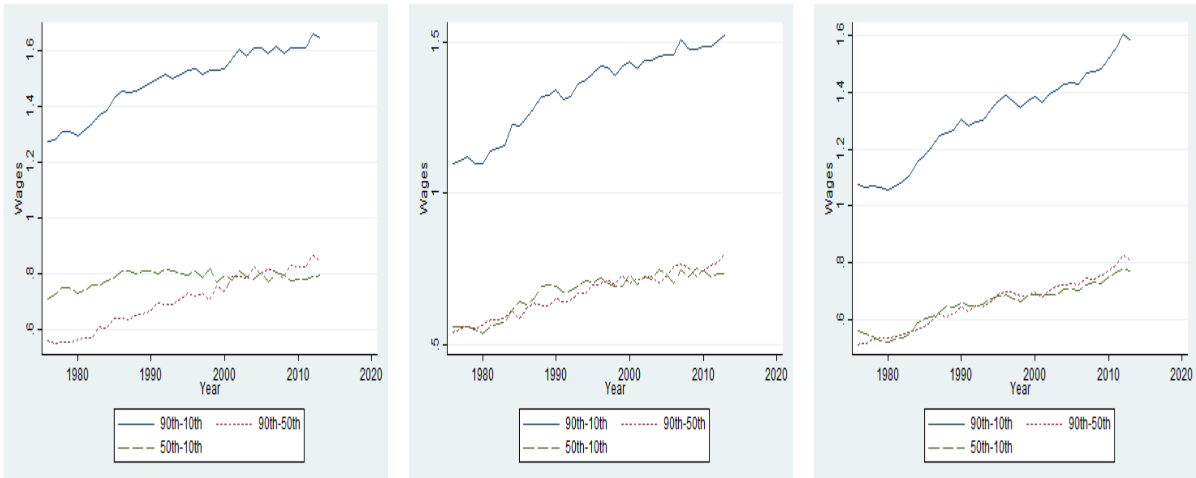


(d) 1976 vs 2006 (e) 1976 vs 2006 (f) 1976 vs 2013

Panel B: Without Correcting for Selection

Figure 5: CDF Comparisons of Women's Wage Distributions Over Time

mid-1990s. Afterward, the series for women continued to increase at a faster rate than the series for men. Our measure of the underlying wage inequality (after addressing the selection) for women exhibits a very similar pattern, although at an even faster rate. The implied evolution of wage inequality within and between gender is consistent with the proposed explanation in Mulligan and Rubinstein (2008).



(a) Male

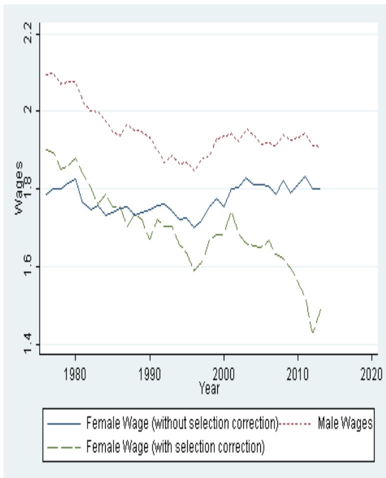
(b) Female (without selection correction)

(c) Female (with selection correction) excluding 2010

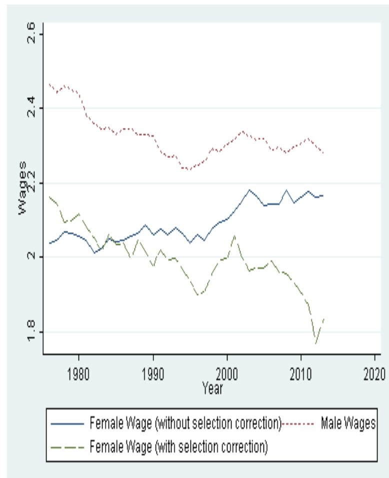
Figure 6: Wage Inequality: $90^{th} - 10^{th}$, $90^{th} - 50^{th}$, and $50^{th} - 10^{th}$ Percentiles of the Distributions

However, further analysis indicates that the sources of the increased inequality among women are completely different between our results and those in Mulligan and Rubinstein (2008). To see this, we plot the time series of select percentiles for both men and women (before and after the selection is controlled for) in Figure (7). It can be seen that the reason for the increase in the *observed* inequality (in their paper) is due to the drastic increase in the 90^{th} percentile of the distribution (without selection correction), while the 10^{th} percentile remained relatively flat during this period. By contrast, using the 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. 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).

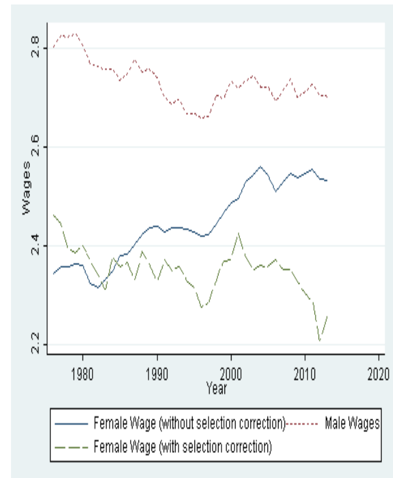
The patterns of the upper-tail (90/50) inequality and lower-tail (50/10) inequality differ across gender. For men there exists a divergence in the time trend of the upper- and lower-tail inequalities. Specifically, the lower tail inequality has increased during the 1980s but stagnated and fluctuated around the same level afterward, while the upper-tail inequality has increased steadily in the past decades. However, the pattern for women is different. Specifically, there has



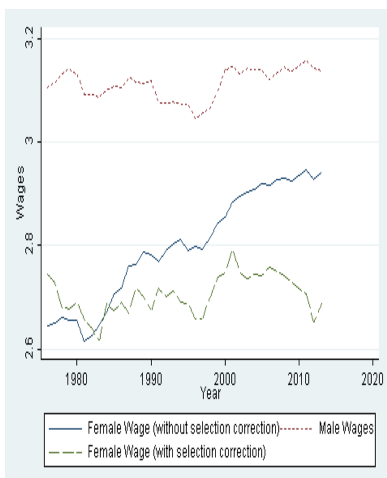
(a) $\tau = 10^{th}$



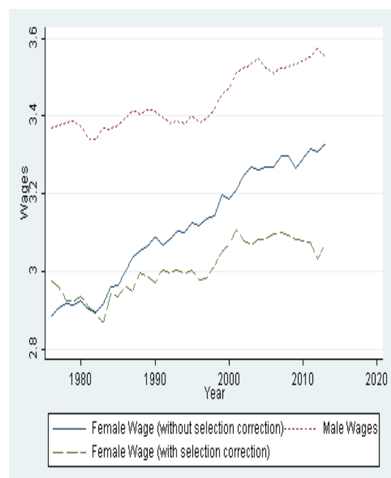
(b) $\tau = 25^{th}$



(c) $\tau = 50^{th}$



(d) $\tau = 75^{th}$



(e) $\tau = 90^{th}$

Figure 7: Wage at Different Parts of the Distributions
Note: Year 2010 is excluded to better reflect the time trend.

been a steady increase in *both* upper- *and* lower-tail inequalities since the mid-1970s. This result holds, whether or not the selection issue is addressed. These observed patterns are consistent with Autor et al. (2008).[19] More important, our results lend further support to their results by showing that the observed pattern for women in their paper is not an artifact of the selection issue.

6.6.2. Explanation 2: Differential Impacts of Expanded Child Care Availability on Women

The second explanation is related to the differential impacts of expanded availability for child care over time on women's employment at different income levels. According to the 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.[20] Meanwhile, the child care costs have also been increasing but affected families at different income levels very differently. According to the census data in 2011, families with employed mothers whose monthly income was 4,500 or more paid roughly 6.7% of their family income (an average of 163 a week) for child care, while those with monthly income of less than 1,500 paid about 40% of their family income (an average of 97 a week).[21] 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.

7. Further Explorations by Educational Groups

In this paper our focus has been to present "the gap" by delving deeper into the question of "what is the gap"? Further work is required to identify sources of the gap. The attribution to different causes will remain as an extensive research agenda. We make a start in this direction by examining our findings decomposed by educational groups. Blau (1998) examines the trends in the well-being of American women by educational groups during the period 1970-1995. One of her findings is "the deteriorating relative economic position of less educated women." To this end, we further explore the gender gap changes across educational groups. We repeat our analysis for four different educational groups (below high school education, high school, some college, and college and above). The results are reported in Table (6). In the interest of space, the results for the conventional measures of the gender gap are reported in the supplemental material. We also thank Jim Heckman for the suggestion that eventually leads to these further results.

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

[20]Source: `http://www.census.gov/hhes/childcare/data/service/facilities/`

[21]Source: `http://www.pewresearch.org/fact-tank/2014/04/08/rising-cost-of-child-care-may-help-explain-in`

Table 6: Measures of The Entropy Gender Gap by Educational Groups

	Without Selection Correction				With Selection Correction			
Year	Less Than High-School (1)	High School (2)	Some College (3)	College & Above (4)	Less Than High-School (5)	High School (6)	Some College (7)	College & Above (8)
1976	16.11	13.71	10.53	9.44	10.35	8.35	8.22	9.49
1977	14.34	12.82	10.70	9.14	12.42	9.54	9.88	10.29
1978	13.62	12.64	10.24	9.90	11.90	12.24	11.14	12.47
1979	12.36	12.11	10.57	10.00	12.93	11.63	11.41	11.64
1980	13.41	12.22	10.46	9.24	11.60	11.26	10.55	10.43
1981	11.56	11.25	9.70	9.19	10.93	10.59	10.75	10.16
1982	9.81	10.64	9.93	9.21	10.13	10.40	11.14	11.47
1983	11.81	9.45	8.09	8.22	15.17	10.89	11.00	11.05
1984	9.60	8.06	7.65	8.01	9.36	7.61	8.76	9.28
1985	8.45	7.62	6.89	7.97	9.66	8.65	9.03	10.31
1986	8.16	6.41	7.58	7.13	9.39	6.89	9.77	9.29
1987	6.74	6.19	6.52	5.92	10.44	8.79	10.19	9.71
1988	7.10	5.49	4.84	6.28	10.06	7.10	7.42	9.03
1989	6.51	5.84	4.53	5.54	9.88	8.51	8.03	9.31
1990	5.04	4.85	4.82	4.82	10.01	8.69	8.89	8.61
1991	5.03	4.07	4.61	4.59	7.19	5.91	7.55	7.34
1992	6.05	4.03	3.67	4.06	9.13	6.35	7.10	6.83
1993	4.95	3.46	3.29	4.06	8.17	5.93	6.44	6.91
1994	3.89	3.25	2.87	3.37	6.26	6.15	6.10	6.13
1995	4.88	3.08	2.82	3.09	8.69	6.01	6.27	6.10
1996	2.88	3.17	2.90	3.71	8.24	7.51	7.00	7.47
1997	2.84	3.70	2.92	3.15	8.26	8.20	7.79	6.91
1998	3.41	3.43	3.25	3.01	7.36	7.22	7.59	6.44
1999	4.13	3.26	3.16	3.87	6.85	6.19	6.86	7.29
2000	3.94	3.57	2.96	3.31	6.15	6.87	7.36	6.79
2001	3.16	3.13	3.05	3.30	4.97	5.83	6.78	6.69
2002	2.46	2.85	3.09	3.18	7.03	6.87	8.05	7.59
2003	3.57	2.41	2.91	3.28	9.97	7.24	8.90	7.89
2004	3.26	2.32	2.35	2.67	9.12	7.09	8.35	7.06
2005	2.93	2.40	2.63	3.00	8.40	7.39	8.12	7.71
2006	4.58	2.50	2.31	3.10	8.28	6.39	6.79	7.17
2007	3.31	2.35	1.92	2.69	7.86	7.06	6.73	7.14
2008	3.04	2.25	2.18	2.80	9.09	7.66	7.61	7.53
2009	3.43	2.55	2.34	3.14	9.50	8.31	8.01	7.84
2010	3.93	2.60	2.77	2.54	43.77	37.69	35.42	42.19
2011	3.91	2.37	2.16	2.46	13.29	10.79	9.41	7.60
2012	3.30	2.42	2.19	2.61	14.85	13.17	11.72	8.90
2013	2.56	2.28	2.13	2.31	11.04	10.73	9.61	7.29

[1] Data Source: IPUMS CPS (`http://cps.ipums.org/cps/`). Columns (1)-(4) report the entropy gender gap ($\times 100$) without selection correction by educational groups (below high school education, high school, some college, and college and above). Columns (5)-(8) report the results with selection correction.

We first look at the results without addressing selection in Columns (1)-(4) of Table (6). The findings are generally similar across educational groups. The gender gap within each group decreased notably before the mid-1990s, and continued to decrease afterward but at a somewhat modest rate. The starting levels of the gender gap and the implied rates of decline are different across groups. We find that until the early 1990s, the declines in the gender gap are concentrated among workers with high-school and some college education, whereas the declines are relatively smaller for workers with below high school education, and smallest for workers with more than 4-year college education. Over this period, the average annual percentage changes are -7.5 and -6.1 percents for workers with high-school and some college education, respectively. The average annual percentage changes are -5.9 percent for workers with less than high school education and only -4.6 percent for those with more than 4-year college education. The gender gap among the least educated workers were larger than the gender gap in the rest of the population. Afterward, the progress among the least educated workers completely stagnated, while women in other groups continued to narrow the gender gap with their male counterparts. Specifically, the average annual percentage changes were only -0.4 percent, while the average annual percentage changes were about -2 percent for college graduates, the largest among all the groups. This pattern is very similar to what is found in Blau (1998). The SD results (reported in the supplemental material) indicate first-order dominance relations for almost all years, except four cases where we fail to observe any dominance relations among the least educated workers. This result again confirms that, without controlling for the selection issue, women in all educational groups fare worse compared to their male counterparts, despite the convergence in the past decades.

As noted in Blau (1998), one interesting question is whether the declining relative wages of the least educated is simply a result of compositional changes within the group. To shed light on this issue, we turn to our results when the selection issue is addressed. We continue to find that the gender gap is larger among the least educated workers than the gender gap among other educational groups. More strikingly, the gender gap among the least educated workers does not exhibit a clear convergence over time, and fluctuates more than other groups. Before 1994, the average rates of changes are actually positive among the least educated workers, while the average annual percentage changes are 0.4, 0.7, and about 1.4 percent for high school graduates, some college, and college graduates, respectively. Our results indicate that "the deteriorating relative economic position of less educated women" over this period could be well underestimated. Moreover, consistent with Blau (1998), our results imply that the compositional changes are responsible for the patterns observed for the least educated.

Several other findings are worth noting. First, the gender gap was stagnant between mid-1990s and early 2000s across all educational groups, and more strikingly, it started to increase afterward. The average rates of increase are largest for college graduates (18.7 percent over the period of 1994 and 2013), while they are similar for other groups. Second, turning to our SD

Table 7: Stochastic Dominance Results (Male vs Female Potential Wage Distributions corrected for selection) by Education

	Less than High School			High School			Some College			College & Above		
Year	SD	d	s	SD	d	s	SD	d	s	SD	d	s
1976	N	0.42	1.33	F	-0.58	-0.93	N	0.68	4.49	N	0.58	4.45
1977	F	-0.29	-0.30	N	0.95	4.00	N	0.49	2.63	N	0.86	5.12
1978	F	**-0.29**	**-0.46**	F	-0.57	-0.58	F	-0.41	-0.76	N	0.52	2.3
1979	N	0.60	2.13	F	**-0.89**	**-1.19**	N	0.34	1.29	N	0.53	1.99
1980	N	0.54	2.22	N	0.80	2.94	F	-0.45	-0.46	N	0.64	3.83
1981	F	-0.30	-0.30	N	0.62	0.98	N	0.82	5.36	N	0.98	5.93
1982	N	0.26	0.58	F	-0.62	-0.91	N	0.52	1.52	N	0.75	4.15
1983	F	**-0.26**	**-0.3**	F	**-0.57**	**-0.76**	F	-0.41	-0.67	N	0.71	3.45
1984	N	0.47	1.26	F	**-0.56**	**-0.56**	N	0.66	2.68	N	0.53	2.08
1985	N	0.37	0.95	F	-0.59	-0.68	N	0.49	2	N	0.75	4.18
1986	F	-0.24	-0.31	N	0.56	1.00	N	0.54	2.84	N	0.86	4.78
1987	F	-0.21	-0.54	F	**-0.58**	**-1.37**	F	**-0.4**	**-0.42**	N	0.73	3.07
1988	F	-0.30	-0.30	F	**-0.84**	**-1.24**	N	0.45	1.3	N	0.76	3.3
1989	N	0.23	0.23	F	-0.78	-0.85	F	-0.4	-0.81	N	0.75	3.2
1990	F	**-0.28**	**-0.28**	F	**-0.86**	**-1.19**	F	**-0.41**	**-0.72**	N	0.58	1.95
1991	F	**-0.26**	**-0.35**	F	**-0.62**	**-1.33**	F	-0.41	-0.81	N	0.94	5.19
1992	N	0.29	0.35	F	**-0.75**	**-1.33**	N	0.62	1.84	N	0.96	4.52
1993	N	0.25	0.45	F	**-0.6**	**-1.13**	F	**-0.43**	**-0.43**	N	0.61	3.17
1994	N	0.17	0.04	F	**-0.59**	**-0.89**	F	-0.53	-0.99	N	0.44	1.04
1995	F	**-0.2**	**-0.25**	F	**-0.81**	**-1.14**	F	**-0.48**	**-0.48**	N	0.65	3.74
1996	F	**-0.21**	**-0.22**	F	**-0.68**	**-0.82**	F	**-0.42**	**-0.83**	F	-0.5	-0.64
1997	F	**-0.21**	**-0.4**	F	**-0.64**	**-1.61**	F	**-0.41**	**-1.08**	N	0.74	2.74
1998	F	**-0.19**	**-0.37**	F	**-0.57**	**-0.7**	F	**-0.41**	**-0.76**	N	0.55	1.05
1999	F	**-0.23**	**-0.24**	F	**-0.68**	**-1.38**	F	**-0.58**	**-0.86**	N	0.69	1.96
2000	F	**-0.27**	**-0.34**	F	**-0.54**	**-0.87**	F	-0.43	-0.6	N	0.69	2.24
2001	F	-0.24	-0.34	F	**-0.72**	**-0.86**	F	**-0.69**	**-1.52**	N	1.03	4.18
2002	F	**-0.24**	**-0.34**	F	**-0.81**	**-1.05**	F	**-0.66**	**-1.33**	N	0.6	0.79
2003	F	**-0.27**	**-0.38**	F	**-0.67**	**-1.08**	F	**-0.58**	**-1.02**	N	0.68	1.55
2004	F	**-0.23**	**-0.23**	F	**-0.8**	**-1.08**	F	**-0.65**	**-1.13**	N	0.86	2.93
2005	F	**-0.25**	**-0.39**	F	**-0.7**	**-0.7**	F	**-0.53**	**-1.35**	N	0.64	1.19
2006	F	**-0.27**	**-0.6**	F	**-0.81**	**-0.88**	F	**-0.51**	**-1.14**	N	0.69	1.46
2007	F	**-0.25**	**-0.5**	F	**-0.77**	**-0.77**	F	**-0.53**	**-0.86**	N	0.64	1.61
2008	F	**-0.28**	**-0.39**	F	**-0.65**	**-1.14**	F	**-0.67**	**-0.75**	F	-0.92	-0.92
2009	F	**-0.19**	**-0.38**	F	**-0.66**	**-1.14**	F	**-0.55**	**-0.81**	F	-1.1	-1.1
2010	N	3.75	22.14	N	4.97	22.09	N	5.06	24.72	N	14.39	97.97
2011	F	**-0.22**	**-0.25**	F	**-0.74**	**-1.46**	F	**-0.67**	**-0.76**	F	-0.56	-1.52
2012	F	**-0.2**	**-0.23**	F	**-0.77**	**-1.34**	F	**-0.66**	**-0.91**	F	**-0.58**	**-2.04**
2013	F	**-0.21**	**-0.29**	F	**-0.73**	**-1.1**	F	**-0.52**	**-1.01**	N	0.71	1.25

[1] **N** denotes no (first- or second-order) dominance found, while **F** denotes First-order dominance. Only the d and s statistics are reported. Statistics in bold are significant at 90^{th} percentile.

results (Table 7), we find that the results for individuals with education less than 4-year college (below high school, high school, and some college) are very similar to the full-sample results above. For these individuals, we continue to find that in early years, there are less occurrence of dominance relations in early years than later years (because of the selection pattern). However, we fail to find dominance relations in nearly all years for individuals with more than college education. This result implies that women with college education do not necessarily fare worse than their male counterparts, while women with less education do.

8. A New Concept of the Gender Gap: Based on Alternative Wage Distribution for Women

Our discussions have thus far focused on comparisons of the distributions of men's wage and women's *potential wage offers*. Remember that our purpose of comparing these two distributions is to evaluate women's relative well-being. For those women who do not work full-time, the wage offers may not necessarily represent the wage levels that they actually enjoy. Instead, the actual monetary value of not working is captured by their reservation wages. An interesting and informative comparison would then be based on an alternative wage distribution for women, defined by the mixed distribution of wage offers conditional on full-time employment and reservation wages conditional on non-full-time employment.

We can recover the distribution of reservation wages given unemployment by exploiting the selection equations (14) and (15). This involves a three-step procedure. First, having estimated our selection equation (15), we can obtain $\widehat{g(z)}$, the difference between the potential wage offers and reservation wages. Second, we can then draw potential wages, $\ln(w) = x'\beta(u)$ given non-full-time employment ($S = 0$) for given x and estimated coefficients, $\widehat{\beta}$. Finally, it is straightforward to recover the distribution of reservation wages in the third step. In fact the quantiles of this distribution have a nice expression, which depends on β and the copula of (U, V). In a different context, Bonhomme et al. (2014) rely on the selection equation to recover the distribution of agents' underlying preferences in a similar way. This entire subsection were indeed inspired by our helpful discussions with Stephane Bonhomme, to whom we are grateful.

The S_ρ and SD tests are presented in Table (8).[22] 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 1976, 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. However, the entropy gender gap has been steadily increasing over time. The

[22]Other conventional measures are available in the supplemental material.

time series of our entropy measure is plotted in Figure (8). This result implies that women's relative labor market performance may worsen over time. This result is confirmed by our SD results in Table (8). As above, we again fail to find first-order and even any dominance relations in many cases during the early years, while we observe first-order dominance relations in later years. This is consistent our results above. This again implies that women's relative well-being (measured by wages) may have deteriorated over time.

Table 8: THE **NEW** GENDER GAP MEASURE ($S_\rho \times 100$) AND SD TESTS (BASED ON FEMALE POTENTIAL AND RESERVATION WAGE)

Year	S_ρ (1)	**SD** (2)	d (3)	s (5)	**Year**	S_ρ (5)	**SD** (6)	d (7)	s (8)	**Year**	S_ρ (9)	**SD** (10)	d (11)	s (12)
1976	0.95	N	6.64	15.5	1989	2.80	F	**-0.88**	**-0.88**	2002	4.39	F	**-1.2**	**-1.63**
1977	0.98	N	4.98	6.87	1990	3.11	F	**-0.94**	**-0.94**	2003	4.53	F	**-1.47**	**-2.14**
1978	1.36	S	1.89	-1.52	1991	1.96	F	**-0.94**	**-1.46**	2004	3.87	F	**-1.14**	**-1.15**
1979	1.78	S	1.16	-1.14	1992	2.09	F	**-0.98**	**-0.99**	2005	3.87	F	**-1.35**	**-1.91**
1980	1.98	N	1.06	3.43	1993	2.07	F	**-1.19**	**-1.35**	2006	3.61	F	**-1.63**	**-2.63**
1981	1.72	N	1.35	3.79	1994	2.15	F	**-0.86**	**-0.86**	2007	3.87	F	**-1.09**	**-1.25**
1982	1.29	N	1.32	3.75	1995	2.52	F	**-1.58**	**-1.93**	2008	4.41	F	**-1.62**	**-1.67**
1983	1.12	N	1.04	2.49	1996	3.54	F	**-1.05**	**-1.05**	2009	4.54	F	**-1.12**	**-1.12**
1984	0.95	N	1.38	6.7	1997	3.96	F	**-1.09**	**-1.31**	2010	16.64	N	19.14	119.05
1985	1.37	N	1.27	4.93	1998	3.73	F	**-0.95**	**-0.96**	2011	4.44	F	**-1.01**	**-1.31**
1986	1.44	N	1.21	5.21	1999	3.87	F	**-1.19**	**-1.53**	2012	5.67	F	**-0.98**	**-1.02**
1987	2.19	F	**-0.82**	**-0.82**	2000	4.16	F	**-1.1**	**-1.2**	2013	4.43	F	**-0.96**	**-2.55**
1988	2.14	F	**-1.12**	**-1.23**	2001	3.44	F	**-1.2**	**-2.52**					

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

9. Further Implications for the Literature

9.1. Relationship between Full-time Employment Rate and Selection

Some of the existing literature has suggested that the selection bias may decrease as female employment rates increase. For example, Olivetti and Petrongolo (2008) find that countries with greater gender gap in employment rates (featuring lower female employment rates) are associated with smaller gender gaps. They argue that this is because working women generally have higher wages, and as more women are employed and the selection bias becomes smaller, but the gap increases. Smith and Ward (1989) similarly suggest that the selection bias had become smaller during the 1980s as more women entered the labor force.

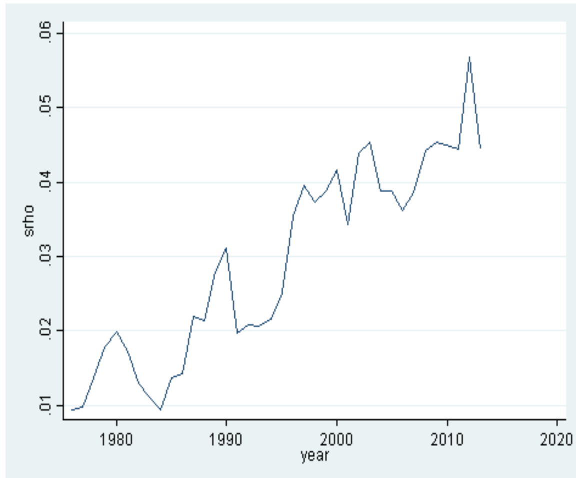


Figure 8: New Gender Gap Measure S_ρ: Male vs Female Wage (Potential and Reservation Wage) Distributions
Note: Year 2010 is excluded to better reflect the time trend.

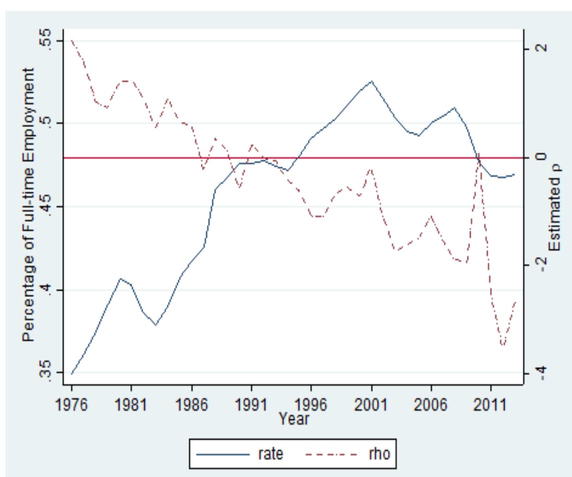


Figure 9: Percentage of Full-time Female Workers and Absolute Value of the Estimated ρ (measuring selection)

As pointed out in Mulligan and Rubinstein (2008), such argument is based on the assumption of a fixed, positive selection. Our results above have already indicated that the selection rules may change over time. Given the estimated ρ, we can also get a direct sense of how the direction and magnitude of the selection varies over time with the employment rates. We plot such relationship in Figure (9). The solid line is the full-time employment rates among women, and the dashed line is the estimated ρ. As we can see, there exists a negative selection when the employment rates are extremely low, while there exists a positive selection when the employment rates are relatively higher. Interestingly, the magnitude of the selection bias does get smaller as the employment rates rise, although the relationship is not necessarily monotonic. For example, the size of the selection parameter (negative selection) continued to decline with the increase in the employment rates before 1992 and became roughly zero when the full-time employment rate reached roughly 50% in 1992. Afterward, the magnitude of the selection parameter (positive selection) closely fluctuate with the employment rates. When the employment rates increase, the selection parameter gets closer to the zero horizontal line (i.e., no selection).

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

Recall that the assumption imposed in Blundell et al. (2007) to identify the *true* gender gap is the positive selection into work, which can be expressed as “first-order stochastic dominance of the distribution of wages of nonworkers 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. We formally test this assumption using our results. These results are presented in Table (9), and the corresponding comparisons of the CDFs are in the supplemental material. As we can see, 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 the 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 result does directly challenge the applicability of this assumption at least in the U.S..

9.3. Implications for Decomposition and Potential Policies

Decomposition/counterfactual analysis is informative about potential sources of the gender gap. Such analysis constructs the counterfactual wage distribution when either wage “structure”

Table 9: 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	F	**-0.55**	**-0.55**	1989	N	1.28	43.77	2002	F	**-0.91**	**-0.96**
1977	S	0.69	**-0.61**	1990	F	-0.75	-0.76	2003	F	**-0.5**	**-0.91**
1978	S	1.01	**-0.62**	1991	N	1.16	28.69	2004	F	**-0.89**	**-0.96**
1979	S	1.06	**-0.8**	1992	N	1.12	21.75	2005	F	**-0.88**	**-0.94**
1980	S	1.05	**-0.7**	1993	N	0.93	17.54	2006	F	**-0.89**	**-0.9**
1981	S	1.02	**-0.71**	1994	F	**-0.79**	**-0.79**	2007	F	**-0.92**	**-0.99**
1982	S	1.2	**-0.68**	1995	F	**-0.74**	**-0.78**	2008	F	**-0.95**	**-0.96**
1983	N	1.92	99.12	1996	F	**-0.67**	**-0.69**	2009	F	**-0.62**	**-0.89**
1984	S	1.6	**-0.87**	1997	F	**-0.71**	**-0.73**	2010	N	14.86	901.37
1985	N	1.86	100.03	1998	F	**-0.32**	**-0.73**	2011	F	**-0.89**	**-0.92**
1986	N	1.31	64.2	1999	F	**-0.74**	**-0.84**	2012	F	**-0.88**	**-0.88**
1987	N	0.94	11.71	2000	F	**-0.74**	**-0.75**	2013	F	**-0.86**	**-0.88**
1988	N	1.4	51.25	2001	N	1.13	10.94				

[1] **N** denotes no (first- or second-order) dominance found, while **F** denotes First-order dominance **S** 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 extended working paper version.

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 for an excellent account of this issue). However, most of this type of analysis usually ignores the selection issue.

We also share a similar view with Aaberge et al. (2013) and Carneiro et al. (2001) that eventually there is a need to go beyond decomposing the gender gap and simply obtaining the counterfactual effects at select quantiles. Decomposition could greatly enrich the policy relevance of counterfactual analysis. Structural and compositions effects could inform different types of policies. Structural effects could be closely related to 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 inform us of the potential impact on wages policies aimed at changing women's human capital characteristics such as education. Implementing these policies could potentially result in both "winners" and "losers"; in other words, 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.

The quantile-copula approach can be extended to further recover the counterfactual distribution that corrects for selection. As noted in Chernozhukov et al. (2013), there are two ways of obtaining counterfactual distributions and conducting decomposition analysis: one is based on conditional quantile regression, and the other re-weighting using propensity scores. These two approaches are asymptotically equivalent. However, the second approach is essentially a nonparametric approach, and as noted in Newey (2007) and Huber (2014), without imposing further assumptions, it is impossible to recover the distributions for other groups than the selected subpopulation (i.e., full-time workers in our case). Moreover, in the presence of selection, the counterfactual distributions are also not straightforward to obtain in the re-weighting approach, even for the selected population. We thus adopt the first approach, which can be easily adapted in our context because we are able to estimate the conditional quantile selection regressions.

We do want to emphasize that our focus is on the general welfare implications behind the decomposition analysis that is often overlooked in the literature. Such implications are not unique to a particular decomposition method, but we choose a particular method, which is well studied and can readily accommodate the selection issue, to illustrate our proposals.[23]

[23] In a companion paper, we employ the re-weighting approach. 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.

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.[24] The counterfactual distribution can be recovered as follows,

$$F_{Y_c}(y) = F_{Y\langle i|j\rangle}(y) = \int F_{Y_i|X_i}(y|x)dF_{X_j}(x) \tag{16}$$

From Equation (10), it follows that Equation (16) can be re-written as follows

$$\begin{aligned} F_{Y_c}(y) = F_{Y\langle i|j\rangle}(y) &= \int\{\int_0^1 I[Q_\tau(Y_i|X_i) \leq y]d\tau\}pdF_{X_j}(x) && (17) \\ &= \int\{\int_0^1 I[X\beta_i \leq y]d\tau\}dF_{X_j}(x) && (18) \end{aligned}$$

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

$$\begin{aligned} F_{Y_{c1}}(y) &= \int\{\int_0^1 I[X\beta_m \leq y]d\tau\}dF_{X_f}(x) \quad \text{(Counterfactual Distribution \#1)} && (19) \\ F_{Y_{c2}}(y) &= \int\{\int_0^1 I[X\beta_f \leq y]d\tau\}dF_{X_m}(x) \quad \text{(Counterfactual Distribution \#2)} && (20) \end{aligned}$$

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".

The entropy gender gap is presented in Table (10). First, regardless of whether we control for the selection, the structural effects appear to be much greater than the composition effects. The latter is rather small and often close to zero. Second, failure to control for selection often underestimates the structural effects for females, but overestimates the composition effects. Note that in the case of 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.

Turning to SD tests in Table (11). We first note that regardless of whether we control for the selection, we observe that the female wage distributions with male wage structure first-order dominate the female wage distribution. This result implies that, in these cases, changing

[24] 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.

Table 10: The Entropy Gap between the Female and Counterfactual distributions

	Female vs Counterfactual Distributions													
	#1 (Strucutural)		#2 (Composition)			#1		#2			#1		#2	
	Selection Correction					Selection Correction					Selection Correction			
Year	Yes	No	Yes	No	Year	Yes	No	Yes	No	Year	Yes	No	Yes	No
	(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)		(9)	(10)	(11)	(12)
1976	5.48	8.43	0.20	0.43	1989	5.02	3.49	0.14	0.48	2002	3.98	2.13	0.05	0.51
1977	6.62	7.79	0.16	0.48	1990	5.19	3.12	0.13	0.35	2003	4.67	1.94	0.05	0.44
1978	8.40	7.93	0.14	0.40	1991	3.49	2.73	0.11	0.39	2004	4.36	1.73	0.04	0.51
1979	8.43	8.18	0.13	0.47	1992	3.54	2.49	0.11	0.39	2005	4.31	1.95	0.03	0.48
1980	7.27	7.83	0.12	0.41	1993	3.26	2.38	0.13	0.38	2006	3.95	1.99	0.03	0.50
1981	6.74	7.36	0.14	0.35	1994	3.30	2.16	0.13	0.32	2007	4.43	1.93	0.02	0.55
1982	7.23	7.32	0.17	0.38	1995	3.66	2.29	0.11	0.33	2008	4.74	1.89	0.03	0.52
1983	7.11	6.03	0.22	0.43	1996	4.19	2.19	0.10	0.30	2009	4.84	2.05	0.05	0.51
1984	5.08	5.42	0.20	0.43	1997	4.43	2.33	0.09	0.50	2010	2.18	2.08	0.08	0.52
1985	5.64	4.97	0.18	0.33	1998	3.62	2.23	0.06	0.44	2011	5.60	1.92	0.14	0.43
1986	5.12	4.48	0.19	0.32	1999	3.72	2.43	0.06	0.41	2012	6.48	1.95	0.16	0.48
1987	5.88	4.03	0.20	0.36	2000	3.54	2.03	0.06	0.46	2013	5.52	1.72	0.14	0.47
1988	4.58	3.67	0.14	0.37	2001	3.12	2.12	0.04	0.45					

Notes: Counterfactual distribution #1 (based on male wage structure); counterfacutal distributions #2 (based on male characteristics).

earnings structure would result in a change in the earnings distribution for women, and that the change is *uniformly* in favor of all women. 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). However, our results are much stronger. Recall that the existing literature typically relies on regression analysis at the mean.

When we do not control for selection, we observe second-order dominance relations only in early years, but no meaningful SD ranking of the female wage distribution and the counterfactual wage distribution (#2) in later years. This result implies that even if dispersion is incorporated into the welfare criteria, changing the distribution of women's human capital characteristics to the distribution of men's characteristics may not necessarily represent welfare improvement. An implication is that the *distribution* of women's human capital characteristics is not necessarily inferior to that of men's, and thus that policies aimed at changing the human capital characteristics only, instead of wage structure, may not lead to welfare improvement for women. However, once the selection issue is controlled for, we find that the results are drastically different. We instead observe first-order dominance relations in every year. These results are consistent with the selection pattern found above. The observed samples in later years consist of high-wage workers whose human characteristics are potentially getting close to men's, it is then less likely to observe a dominance relation. However, once *all* women are considered, improving their composition can lead to a better outcome.

These results highlight the potential effectiveness of hypothetical policies aimed at both improving women's wage structure and their characteristics. They also indicate potentially misleading policy conclusions with failure to account for selection, especially in regards to composition effects.

Certain assumptions underly this type of analysis. As noted in Fortin et al. (2011), the standard assumption used in this type of counterfactual analysis 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)).[25] 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.[26]

10. Conclusions

We find that aggregation of the gap at all quantiles matters and is challenging. It is always subjective and dominance tests provide complementary guidance to the degree of challenge in subjectivity, as well as invite a closer look at separate quantiles. We find selection is a significant issue, as is heterogeneity.

Our approach and methods have implications and utility for the analysis of a range of topical policy issues. First, our approach can be extended to investigate 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 and many of his followers have argued that "in evaluating well-being, the value-objects are the functionings and capabilities" (Sen 1992, p.46). "The relevant functionings" can include "such elementary things as ... being in good

[25]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."

[26]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 practically infeasible given the number of variables that people typically include in the wage equations.

Table 11: Stochastic Dominance Results (Female Wage Distribution and the Counterfactual distribution #1 & #2)

	Female vs. Counterfactual Distribution											
	#1 (Structural Effects)						#2 (Composition Effects)					
	Section Correction						Section Correction					
	Yes			No			Yes			No		
Year	SD	d	s	SD	d	s	SD	d	s	SD	d	s
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1976	F	-6.57	-6.57	F	-0.54	-0.69	F	-6.06	-6.90	S	0.80	-0.69
1977	F	-7.32	-7.32	F	-0.64	-0.67	F	-6.17	-6.17	S	1.16	-0.74
1978	F	-9.44	-12.22	F	-0.95	-0.95	F	-6.20	-8.12	S	0.99	-0.74
1979	F	-10.56	-10.79	F	-0.95	-1.01	F	-6.74	-7.66	S	1.14	-0.89
1980	F	-8.39	-8.39	F	-0.75	-0.75	F	-7.00	-8.71	N	1.71	8.33
1981	F	-11.94	-11.94	F	-0.73	-1.02	F	-6.60	-6.60	N	1.65	2.37
1982	F	-8.42	-8.54	F	-0.80	-0.80	F	-6.42	-7.73	S	1.37	-0.83
1983	F	-7.54	-7.54	F	-0.81	-0.81	F	-7.09	-7.09	N	1.44	0.40
1984	F	-7.84	-8.45	F	-0.94	-1.14	F	-6.35	-6.35	N	1.74	0.66
1985	F	-8.26	-8.26	F	-0.72	-0.94	F	-6.85	-7.66	N	1.56	2.77
1986	F	-8.18	-8.18	F	-0.76	-0.85	F	-6.25	-6.25	S	1.36	-0.65
1987	F	-7.56	-11.15	F	-0.77	-0.89	F	-7.32	-8.03	S	1.47	-0.79
1988	F	-11.01	-11.01	F	-0.83	-0.88	F	-7.99	-7.99	N	1.51	9.40
1989	F	-9.16	-13.35	F	-0.77	-0.77	F	-6.41	-6.41	N	1.49	8.58
1990	F	-7.69	-7.76	F	-0.83	-0.88	F	-7.73	-7.73	N	1.62	8.57
1991	F	-10.95	-11.54	F	-1.28	-1.28	F	-7.14	-8.23	N	1.38	5.54
1992	F	-7.79	-14.18	F	-0.93	-0.93	F	-6.78	-6.78	N	1.37	5.25
1993	F	-7.70	-11.26	F	-0.88	-0.88	F	-6.53	-6.71	N	1.39	6.32
1994	F	-9.81	-10.32	F	-0.73	-0.73	F	-7.79	-7.79	N	1.12	4.28
1995	F	-9.62	-9.84	F	-0.83	-0.83	F	-6.65	-6.65	N	1.11	3.74
1996	F	-7.68	-10.61	F	-0.96	-1.44	F	-6.42	-6.42	N	1.05	2.47
1997	F	-8.04	-10.95	F	-0.76	-0.76	F	-6.39	-7.22	N	1.01	5.06
1998	F	-7.90	-7.90	F	-0.77	-0.77	F	-6.88	-7.30	N	0.93	2.44
1999	F	-7.44	-9.45	F	-0.79	-1.03	F	-6.63	-6.88	N	1.00	2.30
2000	F	-7.22	-7.91	F	-0.75	-0.75	F	-7.17	-7.17	N	1.26	4.42
2001	F	-10.98	-14.59	F	-1.13	-1.68	F	-9.16	-9.16	N	1.45	4.86
2002	F	-11.24	-11.24	F	-1.02	-1.15	F	-8.73	-9.98	N	1.31	3.70
2003	F	-9.68	-14.10	F	-1.03	-1.85	F	-8.25	-8.25	N	1.42	3.32
2004	F	-10.14	-10.14	F	-1.09	-1.09	F	-8.16	-8.22	N	1.26	5.07
2005	F	-9.58	-10.69	F	-1.07	-1.07	F	-7.90	-8.32	N	1.17	3.34
2006	F	-9.29	-14.23	F	-1.04	-1.67	F	-8.37	-8.44	N	1.60	4.47
2007	F	-9.08	-9.96	F	-0.98	-1.20	F	-8.04	-8.04	N	0.93	2.93
2008	F	-8.81	-14.10	F	-0.99	-0.99	F	-7.75	-7.75	N	1.25	2.59
2009	F	-10.57	-12.40	F	-1.09	-1.36	F	-8.24	-9.36	N	1.46	2.77
2010	F	-10.91	-12.83	F	-1.06	-1.52	F	-7.53	-7.78	N	1.44	3.53
2011	F	-11.21	-14.20	F	-1.11	-1.35	F	-7.98	-9.83	N	1.16	2.07
2012	F	-8.96	-11.44	F	-1.04	-1.04	F	-7.60	-8.72	N	1.09	3.10
2013	F	-8.76	-10.14	F	-1.06	-1.06	F	-8.66	-9.39	N	1.25	2.25

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

health" (Sen 1992, p.39). Blau (1998) demonstrates the importance of using a broad range of indicators of well-being in "forming a more complete picture of changes in women's well-being than may be obtained elsewhere in the literature." Gender differences in such variables as wages, occupations, and leisure are considered in her paper, separately. As noted in Wu et al. (2008), "examining each attribute separately can sometimes lead to misleading welfare inferences". While a growing literature has developed investigating multi-dimensional welfare measures that take into account earnings and other factors jointly (e.g. Wu et al., 2008), there has not been any development for the gender gap (to the best of our knowledge). Our approach could be particularly useful because both entropy measure and SD analysis are constructed over the space of *distributions* and can be seamlessly applied to univariate and multi-outcome contexts. Moreover, the approach to address the selection issue may be similarly applied to obtain these distributions when it is indeed a concern.[27]

Second, while in this paper we focus on measurement and analysis of the broad gender earnings gap, our approach is readily adapted to measure and analyze other types of earnings distances, such as advantaged vs. disadvantaged groups (e.g. white v.s. black – the racial gap). All the issues and discussions are similarly relevant in these contexts. For example, in terms of the selection issue, Neal (2004) find that the *median* racial gap may be severely underestimated because women's labor force participation decisions differ across races, but we could further extend this result to the distributional level as well as provide more rigorous comparison of the welfare of blacks and whites.

Third, our paper is also related to the broader inequality literature. In addition to between-group inequality (the gender gap), we are also able to speak to the evolution of within-group inequality across gender (which is also part of the explanation for our results). Specifically, being able to recover the entire wage offer distribution for women, we can directly assess how the overall inequality (captured by 90/10 log wage differential), upper-tail (captured by 90/50 log wage differential) and lower tail (captured by 50/10 log wage differential) inequalities change over time for women. Our results lend further support to the results found in Juhn et al. (1993) and Autor et al. (2008) by showing that their results for women are not simply an artifact of the selection issue.

Forth, a less obvious implication of our proposals is for the literature on program evaluation. Instead of comparing the outcomes between men and women, this area is generally concerned with estimation of treatment effects, differences in the raw and counterfactual outcomes (the former group is the outcome in the presence of a policy/program intervention, while the latter is the outcome in the absence of such policy). It is conceivable that implementation of these

[27]For example, women and men could have different survival rates in developing countries. If health is an attribute of interest, comparison based on the sample that we observe and hence those who survive could be similarly misleading.

policies could result in both "winners" and "losers" – some people may gain from the policy while others may lose. This in turn translates into both positive and negative (quantile) treatment effects. This is indeed the case where a more rigorous welfare assessment will become useful. Our examination of the entropy measure and SD analysis could be easily adapted in this case. Moreover, the literature on program evaluation also often ignores the selection issue. Here we illustrate this implication using a general decomposition analysis.

Finally, aggregate time series of the gender gap can be used for further empirical analysis. For example, Biddle and Hamermesh (2011) has noted that little is known about how wage differentials vary with the extent of labor market conditions. Our measure can directly be used for this purpose (in the spirit of Ashenfelter (1970)) to examine the aggregate relationship between the gender gap and the aggregate unemployment rate.

The work of attribution of the gap to separate covariates and sources remains as a major undertaking. We make a start by highlighting the differences by educational attainment, as well as conducting decomposition analysis.

11. 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.

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

Arellano, M. and S. Bonhomme. 2012. "Quantile Selection Models: with an Application to Understanding Changes in Wage Inequality." *Unpublished Manuscript* .

Ashenfelter, O. 1970. "Changes in Labor Market Discrimination Over Time." *Journal of Human Resources* 5:403–30.

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.

Balasubramanian, V., J.J. Heckman, and A. Maloney. 2014. "Relative Entropy and Proximity of Quantum Field Theories." *arXiv:1410.6809 [hep-th]* .

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.

Biddle, J. and D.S. Hamermesh. 2011. "Cycles of Wage Discrimination." *NBER Working Paper* 17326.

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 A.H. Beller. 1988. "Trends in Earnings Differentials By Gender, 1971-1981." *Industrial and Labor Relations Revew* 41:513–529.

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

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

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

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

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, ChristopherR. and BarryT. 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.

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

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

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

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

Firpo, S., N. Fortin, and T. Lemieux. 2007. "Decomposing Wage Distributions using Recentered Influence Function Regressions." *Unpublished Manscript* .

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

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

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

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.

Hart, J. 1997. *Nonparametric Smoothing and Lack-of-Fit Tests*. New York: Springer-Verlag.

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

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

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

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

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

Hirsch, BarryT. and EdwardJ. 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. 2011. "Testing Instrument Validity in Sample Selection Models." *University of St. Gallen, Discussion Paper* 2011-45.

Huber, M. and B. Melly. 2012. "A Test of the Conditional Independence Assumption in Sample Selection Models." *Unpublished Manuscript* .

Imbens, G.W. 2004. "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *Review of Economics and Statistics* 86:4–29.

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.

King, M., S. Ruggles, A.J. Trent Alexander, S. Flood, K. Genadek, M.B. Schroeder, B. Trampe, and R. Vick. 2010. *Integrated public use microdata series, current population survey: Version 3.0. [Machine-readable database]*. University of Minnesota, Minneapolis, Minnesota.

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

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

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. Racine. 2007. *Nonparametric Econometrics*. Princeton, New Jersey: Princeton University Press.

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

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

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

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

Maasoumi, E. 1999. *Handbook of Income Inequality Measurement*, chapter Multidimensioned Approaches to Welfare Analysis. New York, NY: Kluwer Academic Publishers.

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

Maasoumi, E., E.B. Dagum, and S. Giannerini. forthcoming. "A Powerful Entropy Test for Linearity Against Nonlinearity in Time Series." *Biometrica* .

Maasoumi, E. and A. Heshmati. 2000. "Stochastic Dominance Amongst Swedish Income Distributions." *Econometric Reviews* 19:287–320.

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

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

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

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

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

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

Newey, W.K. 2007. "Nonparametric Countinuous/Discrete Choice Models." *International Economic Review* 48:1429–1439.

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

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

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

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

Sen, Amartya. 1974. "Informational bases of alternative welfare approaches: Aggregation and income distribution." *Journal of Public Economics* 3:387–403.

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

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

Shorrocks, A.F. 1984. "Inequality Decomposition by Population Subgroups." *Econometrica* 52:1369–85.

Skaug, H.J. and D. Tjostheim. 1996. *Time Series Analysis in Memory of E.J. Hannan*, volume II, chapter Measures of distance between densities with application to testing for serial independence, p. 363?7. New York: Springer.

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.

Wu, X., A. Savvides, and T. Stengos. 2008. "The Global Joint Distribution of Income and Health." *Unpublished Manuscript* .

Yaari, Menahem E. 1987. "The Dual Theory of Choice under Risk." *Econometrica* 55:95–115.

Yaari, Menahem E. 1988. "A controversial proposal concerning inequality measurement." *Journal of Economic Theory* 44:381–397.

Zoli, Claudio. 1999. "Intersecting generalized Lorenz curves and the Gini index." *Social Choice and Welfare* 16:183–196.

Table A1: The Patterns of Changes in Measures of The Gender Gap

Year	$S_\rho \times 100$ (1)	Mean (2)	10th (3)	25th (4)	50th (5)	75th (6)	90th (7)
1977	D	D	D	D	I	I	D
1978	I	I	D	D	D	I	D
1979	D	D	D	D	I	I	I
1980	D	D	D	D	D	D	D
1981	D	D	I	D	I	I	D
1982	D	D	D	I	I	D	I
1983	D	D	D	D	D	D	I
1984	D	D	I	D	D	D	D
1985	D	D	D	D	D	D	I
1986	D	D	D	I	I	D	D
1987	D	D	I	D	I	D	D
1988	D	D	I	D	D	D	D
1989	D	D	D	D	D	D	D
1990	D	D	D	I	D	I	D
1991	D	D	D	D	D	D	I
1992	D	D	D	I	D	D	D
1993	D	D	I	D	I	D	D
1994	D	D	D	D	D	D	D
1995	D	I	I	I	I	I	D
1996	I	I	I	D	I	D	D
1997	D	D	I	I	I	I	D
1998	I	I	D	I	I	D	I
1999	I	I	I	D	D	I	D
2000	D	D	I	I	I	I	I
2001	D	D	D	D	D	D	I
2002	D	D	D	D	D	D	D
2003	D	D	I	D	D	I	D
2004	D	D	I	I	D	D	I
2005	I	I	D	I	I	D	D
2006	D	I	I	D	I	D	D
2007	D	D	I	I	I	I	D
2008	I	D	D	D	I	I	I
2009	I	I	I	I	D	D	I
2010	I	I	D	D	I	I	D
2011	D	D	D	I	I	D	D
2012	I	I	–	D	D	I	I
2013	D	D	D	D	I	D	D

[1] Column (1) reports the entropy gender gap (×100). Columns (2)- (6) report conventional measures.

Table A2: The Patterns of Changes in Measures of The Gender Gap (Corrected for Selection)

Year	$S_\rho \times 100$ (1)	Mean (2)	10th (3)	25th (4)	50th (5)	75th (6)	90th (7)
1977	I	I	I	D	I	I	I
1978	I	I	I	I	I	I	I
1979	D	I	I	D	I	I	I
1980	D	D	D	D	D	D	D
1981	D	D	D	D	D	D	D
1982	I	I	I	I	I	I	I
1983	I	I	I	I	I	I	I
1984	D	D	D	D	D	D	D
1985	I	I	I	I	D	I	I
1986	D	D	D	I	I	D	D
1987	I	I	I	I	I	I	I
1988	D	D	D	D	D	D	D
1989	I	I	I	I	I	I	I
1990	D	I	I	I	I	I	I
1991	D	D	D	D	D	D	D
1992	I	I	D	I	I	I	D
1993	D	D	I	D	D	D	D
1994	D	I	I	D	D	I	I
1995	I	I	I	I	I	I	I
1996	I	I	I	I	I	I	I
1997	D	D	I	I	D	I	I
1998	D	D	D	D	D	D	D
1999	D	D	I	D	D	D	I
2000	D	I	I	I	I	I	D
2001	D	D	D	D	D	D	I
2002	I	I	I	I	I	I	I
2003	I	I	I	I	I	I	I
2004	D	D	D	D	D	D	I
2005	D	D	D	I	I	I	D
2006	D	D	D	D	D	D	D
2007	I	I	I	I	I	I	I
2008	I	I	I	D	I	I	I
2009	I	I	I	I	D	I	I
2010	I	D	D	D	D	D	I
2011	D	I	I	I	I	I	D
2012	I	I	I	I	I	I	I
2013	D	D	D	D	D	D	D

[1] Column (1) reports the entropy gender gap (×100). Columns (2)- (6) report conventional measures.