The Gender Earnings Gap: Measurement and Analysis

Esfandiar Maasoumi
Emory University

Le Wang
University of New Hampshire, Harvard University, & IZA

Abstract

When summary measures of latent concepts such as “the gender gap” fail to be adequately representative, one must seek better definitions and measures. This paper presents a set of complementary concepts and measurements of the gender gap that move beyond the traditional summary comparisons of the earnings distributions. In particular, we propose a new concept of “the gender gap” based on the the distance between *entire* distributions with compelling properties: It is free of outlier effects, is capable of representing populations with heterogeneous gaps at different parts of the outcome distributions, and is invariant to increasing transformations. When the gender gap is different or of even different sign at different quantiles, subjective comparisons become inevitable in any summary, cardinal comparisons. In response, we introduce tests based on stochastic dominance to allow for uniform rankings of the earnings distributions between men and women. Using the Current Population Survey data, we first construct a new series on the gender gap from 1976 to 2011 in the United States. We find that traditional “representative” or moment-based measures underestimate a declining trend in “the gender gap“ during this period. More important, these traditional measures do not necessarily reflect the cyclicality of the gender differentials in earnings distributions, and may even lead to false conclusions about how labor market conditions are related to the gender gap at the aggregate level. Second, while we find first-order stochastic dominance in most cases, even for the recent recession where men were hit harder, we also find a few instances where definite conclusions regarding the gender gap cannot be drawn at all or only under more restrictive social evaluation functions. Finally, we conduct full distribution counterfactual analysis which suggests that, in many cases, altering the earnings structure would be more effective in improving women’s welfare (reducing “discrimination“) than would changing human capital characteristics.

† The authors would like to thank seminar participants at Bentley University, Union College, Emory University, and University of New Hampshire for their helpful comments.

1. Introduction

Studying *the gender gap*, generally referencing the earnings differences between men and women, is an important undertaking. It is at the core of social sciences to understand inequality/inequity in a society, as well as the labor market outcomes, and helps shed light on potential policy directions. Policy makers and economists are interested in several questions. One is, How large is the gender gap? Another is, How do women generally fare compared to men in the labor market? We emphasize the question of What is "the gender gap", quantitatively speaking? The answers to these questions are more complex than is implicitly assumed in many of the current responses. They require a careful analysis of the earnings differentials between groups. When summary notions of "the gap", such as average/mean or median earnings differentials fail to be representative of diverse magnitudes and signs at different parts of the earnings distribution, the very notion of "the gap" is itself in need of reflection and innovation. This paper offers some proposals in this regard, and develops the statistical means for implementing them. We offer alternative summary measures to define "the gap", as well as rigorous means of defining and testing rankings between entire distributions.

Conventional wisdom about the gender gap is that women do not fare as well as men do in the labor market. Often cited to support this view is the examination of the earnings differentials between *average or median* men and women (e.g. Polachek, 2006; Blau and Kahn, 2006). The sign of the difference tells us which group fares better in the labor market, and the magnitude is a measure of the severity of the situation. Although useful, the gender differences reported by these summary measures may not represent the gender earnings gap in all parts of the *whole* distribution of earnings. This is especially so when the sign and/or magnitude of the gap is different at different quantiles of the earnings distribution. Researchers are increasingly aware of these issues, and differences at other parts of the earnings distributions (e.g quintiles and percentiles) are also reported in recent years. These different summary statistics greatly improve our understanding of the extent and location of the gender gap, but may still present a bewildering view of the *distribution* of the gap, both at a given point in time, and its evolution over time. Better and more comprehensive summary measures can help.

All summary measures are aggregation devices which assign subjective, explicit and implicit weights to different groups and parts of distributions (high and low earners, for example). This calls for an examination of potential "uniform"rankings that are robust to the subjective weight distributions/welfare functions that underlie summary measures.

To make some of our points more concrete, consider the following numerical examples for a society with only two men (MA and MB) and two women (FA and FB).

Example 1 The difference in earnings between FA and MA (who have similar characteristics other than gender) is $-\$200$, and that between FB and MB is $-\$500$ (also with similar characteristics other than gender). A typical measure, the average difference, suggests that the gender gap is $-\$350$. However, "the average person" here is not "representative". And reporting either $-\$200$ or $-\$500$ would also ignore the other half of the society and consequently would not summarize the situation well. Note that in this example, at least the summary measure as well as the quantile differences are all negative, implying that men fare better than do women in the labor market.

Example 2 The difference in earnings between FA and MA is $-\$200$, but that between FB and MB is now \$200. The average gender gap is \$0! Again, this fails to be representative, since it suggests there is no gender gap at all. Both the $-\$200$ and \$200 strongly misrepresent the rest of the society. Compared to Example 1, only extremely subjective weights would support any ranking. Since these differences are of opposite signs, they suggest different rankings of the earnings distributions between men and women. Any conclusions would be based on an arbitrary weighting scheme.

Example 3 Consider another example similar to the second one but with more information on actual earnings for each individual. MA earns $\$55,000$ and FA $\$54,800$.

MB earns \$1200 and FB $\$1,000$. The difference at both parts of the distribution is \$200, and in one direction. The average gap of 200 would give the same weight whatever the level of earnings. Given the *additional* information on each person's earnings, greater aversion to inequality may give greater weight to the difference in the lower tail, concluding existence of a greater gender gap in favor of men. However, these types of subjective "preferences" are usually not explicitly stated alongside summary measures. The typical utility function in economics is a von Neumann-Morgenstern type function, being increasing and concave. Can we empirically identify situations in which "the gender gap" will be ranked uniformly by all observers subscribing to *any* member of a family of utility functions?

The measurement problem would be even more acute when examining the time trend of the gender gap. The timing of temporal deviations from the long-run trend could vary across different measures, which could in turn lead to, for instance, a confused sense of the impact of business cycles on the gender gap. We first propose a distributional measure of the gender gap based on the normalized Bhattacharay-Matusita-Hellinger entropy measure proposed by Granger et al. (2004). One important feature, among others, of this measure is its ability to summarize the distance between two *entire* distributions, instead of binary differences at different *parts* of the distributions. Another advantage of our measure is its invariance to transformations, generally, and to log earnings transformation, particularly. Second, we employ stochastic dominance (SD) tests to rank the earnings distributions. The SD tests have been widely used to analyze poverty issues and financial outcomes, but not to analyze the gender earnings gap. The advantage of our SD approach is its explicit welfare underpinning, utilization of the entire earnings distributions, and ability to yield uniform rankings of distributions that are robust across a wide class of welfare functions, as well as underlying (and unknown) earnings distributions. Inferring a high dominance relation implies that comparisons based on multiple measures, while supported, are often unnecessary. Moreover, the inability to infer a dominance relation is equally informative, indicating that any ranking must be based on a particular weighting scheme or a specific welfare function. In the latter instances, conclusions are revealed to be highly subjective and may not be conducive to consensus policy-making.

Our methodology here is broadly applicable, including to counterfactual analysis of policy experiments, in which one compares the potential outcome distribution for individuals impacted by a policy with their actual outcomes. Policymakers and economists are often interested in certain policies to bridge the gender gap and improve the well-being of women. These policies can be loosely classified into two groups: (1) policies aimed at changing women's pay structure and (2) policies aimed at changing their observable characteristics that affect their earnings. These types of policies are related to two major reasons that many believe explain the differences in earnings between women and men: differences in wage structure and differences in human capital characteristics. The former is often identified with "discrimination". It is useful to provide an assessment of the changes in the potential earnings among women resulting from any implemented policy. To evaluate a policy, we need to compare the original earnings distribution with the potential earnings distribution resulting from a policy. We employ new developments for identifying counterfactual earnings distributions based on estimated inverse probability weighting methods, see Fortin et al. (2011).

To illustrate our proposals, we utilize the Current Population Survey (CPS) data 1976 - 2011 in the U.S. for our empirical analysis. We reach several conclusions. First, we find that traditional summary measures severely underestimate the declining trend of the gender gap in the U.S. In particular, our entropy gender gap measure implies the gap narrowed at an average annual rate of about 11% during the period 1976-2011, while the largest annual rate recorded by conventional measures (based on the median) is only 5.2% during the same period. It may be helpful to note that, entropy measures are functions of all the moments of a distribution, much as moment generating functions. As such, they gauge the "convergence" between the entire distributions for men and women accordingly. Our measure may thus be seen as a "broad measure". The broad measure of gender gap dropped precipitously before 1990s, but the convergence has drastically slowed down since. Moreover, even though all measures show a consistently declining trend of the gender gap, the timing of temporal deviations from the long-run varies across different measures, which in turn can lead to a false reading of the impact of business cycles on the gender gap. Our measure indicates that our broad measure of the gender gap is relatively insensitive to changes in economic conditions, except around 2001.

Second, comparing the *actual* earnings distributions, we generally observe first-order stochastic dominance to a degree of statistical confidence throughout the sample period, implying that men have generally performed better in the labor market. To our surprise, we find dominance even for the beginning of the recent recessionary period when men were viewed as having been hit harder than women. This conclusion is robust to a wide class of increasing social welfare functions. However, we do find several cases where the relation is only second-order, and one case where no statistically meaningful dominance exists. In these cases, the inference that men fare better than do women is only supported by a narrower class of social welfare functions that are both increasing and "increasingly averse" to inequality (concave). Moreover, it is less likely to find statistically significant first-order dominance relations during the pre-1994 period. These results altogether suggest that women's labor market situation has over time improved relative to men's. Nevertheless, strong evidence of dominance relations in most cases indicates that the improvement is still far from satisfactory; this result casts doubt on the broad effectiveness of the reforms intended to improve women's "relative" labor market outcomes.

Finally, combining the methods proposed here and the recent development in identification of counterfactual analysis, we compare the actual female earnings distribution with two female "counterfactual"distributions: (a) women earning distribution under the earnings structure of men ("discrimination") and, (b) women earnings distribution should they have men's characteristics. The former captures *structural effects* and the latter *composition ef-*

fects. We find that structural effects are generally more important than composition effects. However, the importance of structural effects has declined over time, while that of composition effects has increased. Our SD results indicate that policies aimed at changing women's pay structure are generally effective in improving women's earnings prospects, while policies aimed at changing women's human capital characteristics are not. However, for policies aimed at changing pay structure, we fail to find first-order dominance, even second-order dominance, to a statistical degree of confidence, in a few cases. This implies that there are "winners" and "losers", and broad policy conclusions cannot be drawn without imposing more subjective weights on subgroups through narrower well being functions. We want to emphasize that the two thought experiments conducted here are by no means precisely identified with specific policies. However, they do represent the traits of the policies generally considered for relative improvement in women's labor market outcomes. Working through such exercises using our proposed tools illustrates the importance of these tools in a broader analysis of the gender gap.

While in this paper we focus on measurement and analysis of the broad gender earnings gap, we believe that our research could be further extended along several dimensions. First, our approach could be 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). Second, in this paper we focus on the earnings as the only attribute of welfare. However, researchers have long recognized that welfare involves not only earnings but other attributes such as health, and, as a result, 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). Our approach is constructed over the space of *distributions* and can be seamlessly applied to univariate and multi-outcome contexts. Finally, aggregate time series of our distributional measure of the gender gap, once obtained, 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 rest of the paper is organized as follows. Section 2 presents the empirical methods employed; Section 3 describes the data; Section 4 discusses the results, and Section 5 concludes.

2. Empirical Methodology

2.1. Basic Notations

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 $N = N_0 + N_1$ 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, $\{\ln(w_i^m)\}_{i=1}^{N_0}$ 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); $F_0(y)$ and $f_0(y)$ are similarly defined for $\ln(w^m)$ (i.e. the log of earnings for males). Individual earnings are determined by both observable characteristics X_i and unobservable characteristics ϵ_i via unknown wage structure functions,

$$\ln(w_i^d) = g_d(X_i^d, \epsilon_i^d) \quad d = m, f$$

This specification implies that the gender gap is from three sources: (1) differences in the distributions of observable human capital characteristics X_i^d (e.g. years of schooling); (2) differences in the distributions unobservable human capital characteristics ϵ_i^d (e.g. innate ability); (3) differences in the wage structures, $g_d(\cdot)$. Note that these wage functions are not restrictive and allow for complicated interactions among X_i^d and ϵ_i^d.

2.2. A Distributional Measure of the Gender Earnings Gap

Usually, the gender gap is defined as the difference in certain parts (or functionals) of the earnings distributions between males and females. For example, average gender gap is the difference in the means of the earnings distribution between men and women ($\mathbb{E}[\ln(w_i^m)] - \mathbb{E}[\ln(w_i^f)]$) (where the mean is the first moment of the earnings distribution). The gender gap at a p^{th} quantile is $\ln(w_i^m)^p - \ln(w_i^f)^p$, where the p^{th} quantile of F_0 (the CDF for women's wage distribution) is given by the smallest value $\ln(w_i^f)^p$ such that $F_1(\ln(w_i^f)^p) = p$; $\ln(w_i^m)^p$ is similarly defined for F_1 (the CDF for men's wage distribution). Even though these measures are all functionals of the wages distributions, none of them is able to summarize the information in the *whole* distribution. This problem is particularly acute when the measures differ in terms of magnitudes and sizes across different measures used. Hence, needed is a distributional measure of the gender gap, or a measure of the distances in the earnings distributions between females and males.

Several commonly used information-based entropy measures such as Shannon-Kullback-Leibler are available to measure the information at the distributional level. However, Shannon's entropy measure as well as almost all other entropy measures are not *metric*; these measures violate the triangularity rule and hence cannot be used as a measure of *distance*.

To this end, we use a *metric* entropy measure S_ρ proposed in Granger et al. (2004), which is a normalization of the Bhattacharya-Matusita-Hellinger measure of distance. It is given by

$$S_\rho = \frac{1}{2}\int_{-\infty}^{\infty}(f_1^{\frac{1}{2}} - f_0^{\frac{1}{2}})^2 dy \tag{1}$$

This measure satisfies several desirable properties as a distance metric between entire distributions: (1) it is well defined for both continuous and discrete variables;[1] (2) it is normalized to zero if Y_1 and Y_0 are equal, and lies between 0 and 1, (3) it is a metric and hence a true measure of *distance*, (4) it is invariant under continuous and strictly increasing

[1] Although (1) presumes that the variables are continuous, one can easily adapt this measure to the case of discrete variables, $S_\rho = \frac{1}{2}\sum(p_1^{\frac{1}{2}} - p_0^{\frac{1}{2}})^2$ where p_1 (p_0) is the marginal probability of the random variable Y_1 (Y_0). This generalization allows us to measure the differences in a broader set of outcomes between groups at the distributional level.

transformation $h(\cdot)$ on the underlying variables.[2] Recall that following the literature we utilize the log of earnings as the variable of main interest. Since the log is a strictly increasing function, our measure of the gender gap is the same, whether we use the raw wages or the log of it. Moreover, entropies are defined over the space of *distributions* and are consequently "dimension-less" as it applies to univariate and multivariate contexts. Economists have been increasingly aware of the fact that evaluation of individual well-being is inevitably a multi-attribute exercise (Lugo and Maasoumi, 2008). This feature may become very useful when we consider the *multidimensional* gender gap measure to incorporate attributes other than wages.

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. As a result, in the analysis below, we instead employ bootstrap re-sampling procedure based on 299 replications to obtain critical values of hypothesis testing of $H_0 : S_\rho = 0$.

Our entropy measure of gender gap gives us information on the strong ranking of two wage distributions. However, it does not directly tell us which distribution is (weakly) uniformly

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

"better" relative to large classes of welfare functions. and under what conditions. Below, we explicitly introduce these concepts to rank/compare two distributions.

2.3. Stochastic Dominance

We employ recent tests for Stochastic Dominance (SD) to enable uniform welfare comparisons of the earnings distributions between females and males ($\ln(w^f)$ and $\ln(w^m)$). The SD approach identifies for which class of social welfare functions rankings of the earnings distributions are possible. In this paper, we consider two classes of social welfare functions that are commonly used in economics and finance. Let U_1 denote the class of all *increasing* von Neumann-Morgenstern type social welfare functions u such that welfare is increasing in wages (i.e. $u' \geq 0$), and U_2 the class of social 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. We are interested in the following scenarios:

Case 1 (First Order Dominance):

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

1. $\mathbb{E}[u(\ln(w^m))] \geq \mathbb{E}[u(\ln(w^f))]$ for all $u \in U_1$ with strict inequality for some u;
2. Or, $F_0(y) \leq F_1(y)$ for all y with strict inequality for some y.

Case 2 (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. $\mathbb{E}[u(\ln(w^m))] \geq \mathbb{E}[u(\ln(w^f))]$ for all $u \in U_2$ with strict inequality for some u;
2. Or, $\int_{-\infty}^{y} F_0(t)dt \leq \int_{-\infty}^{y} F_1(t)dt$ for all y with strict inequality for some y.

These two cases imply rankings of the earnings distributions under different conditions. Specifically, if the case 1 holds ($\ln(w^m)$ FSD $\ln(w^f)$), then the earnings distribution among men is "better" than that among women for all policymakers with increasing utility functions in the class U_1 (with strict inequality holding for some welfare function(s) in the class), since the expected social welfare from $\ln(w^m)$ is larger or equal to that from $\ln(w^f)$. Note that $\ln(w^m)$ FSD $\ln(w^f)$ implies that the average male wages are greater than the average female wages. "However, a ranking of the average wages does not imply that one FSD the other; rather, the entire distribution matters" (Mas-Colell et al., 1995, p.196). 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 welfare functions in the class U_2 (with strict inequality holding for some utility function(s) in the class). Note that FSD implies SSD. One immediate advantage of this approach is that our conclusions do not depend on any specific wage distributions and/or weights assigned to subgroups within then population. This approach is thus able to yield uniform rankings of distributions that are robust across a wide class of welfare functions, rendering comparisons based on specific indices unnecessary, but possible and more broadly supported. Higher order SD rankings are based on narrower classes of welfare functions. For instance, Third Order dominance is associated with welfare functions with increasing aversion to inequality which place greater weight on welfare improving transfers at the lower tails of the earnings distribution.

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:

$$d = \sqrt{\frac{N_0 N_1}{N_0 + N_1}} \min \sup [F_1(y) - F_0(y)] \tag{2}$$

$$s = \sqrt{\frac{N_0 N_1}{N_0 + N_1}} \min \sup \int_{-\infty}^{y} [F_1(t) - F_0(t)] dt \tag{3}$$

The test statistics are based on the sample counterparts of d, and s by replacing CDFs with empirical ones; the empirical CDFs are 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; $\widehat{F_0(y)}$ is similarly defined. The underlying distributions of the test statistics are generally unknown and depend on the data. Following the literature (e.g. Maasoumi and Heshmati, 2000; Millimet and Wang, 2006), we use bootstrap techniques for iid samples based on 299 replications to obtain the actual sampling distributions of the test statistics. This approach estimates the probability the statistics falling in any desired interval, as well as indicate where the sample value of the test statistic lies. For instance, 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. All technical details are presented in Appendix 1.

2.4. Counterfactual Distributions

We are often interested in assessing two types of counterfactual situations: First, what if we interchange the wage structure of women with the wage structure of men, holding the distribution of women's human capital characteristics constant? Second, what if we change the distribution of women's human capital characteristics to that of men's, holding the wage structure unchanged? Will these counterfactual distributions be different from the original one? Will these differences necessarily cover any distance between the earnings distributions? Our proposed approaches can be readily applied to answer these counterfactual questions by measuring the distances between the female earnings distribution and the counterfactual distribution, and by ranking them. An important step is to identify the counterfactual distributions of interest. Specifically, we want to identify the following counterfactual outcome distributions:

$$\ln(w_i^{c1}) = g_0(X_{i1}, \epsilon_{i1}) \quad \text{(Counterfactual Outcome \#1)} \tag{4}$$

$$\ln(w_i^{c2}) = g_1(X_{i0}, \epsilon_{i0}) \quad \text{(Counterfactual Outcome \#2)} \tag{5}$$

F_{c1} (f_{c1}) represents the corresponding CDF (PDF) of the counterfactual outcome $\ln(w_i^{c1})$. F_{c2} (f_{c2}) represents the corresponding CDF (PDF) of the counterfactual outcome $\ln(w_i^{c2})$. Notice that the differences in the distributions of F_{c1} and F_1 ($\ln(w_i^{c1})$ v.s. $\ln(w_i^f)$) come from differences in wage structures; the comparisons of these two distributions thus provide insight into potential *discrimination*. On the other hand, the differences in the distributions of F_{c1} and F_1 ($\ln(w_i^{c2})$ v.s. $\ln(w_i^f)$) come solely from differences in the distribution of human capital characteristics; the comparisons thus provide some insight into the gender gap due to *productivity* differences across gender.

As shown in Firpo (2007, Lemma 1), the counterfactual distributions are identified under the following assumptions:

[A1.] Unconfoundedness/Ignorability: Let (D, X, ϵ) have a joint distribution. For all x, ϵ is independent of D conditional on $X = x$, where, as defined above, $D = 1$ for females and $D = 0$ for males; X, ϵ are observable and unobservable human capital characteristics, respectively.

[A2.] Common Support: For all x, $0 < p(x) = \Pr[D = 1|X = x] < 1$.

The counterfactual outcome CDF of $\ln(w_i^{c1})$ is identified and $F_{c1} = \mathbb{E}[\omega_{c1}(D, X) \cdot I[(\ln(w_i) \leq y)]$, where $\omega_{c1}(D, X) = (\frac{p(x)}{1-p(x)}) \cdot (\frac{1-D}{p})$. The counterfactual outcome CDF of $\ln(w_i^{c2})$, F_{c2}, is similarly identified. In practice, the "score"$p(x)$ is estimated by probit or logit. Here we employed probit. p is the unconditional probability over the corresponding characteristics X. Both assumptions (A1) and (A2) are commonly used in the literature. Assumption (A1) implies here that given the values of observable human capital characteristics X, the distribution of unobservable human capital characteristics such as ability is independent of gender. Assumption (A2) rules out the possibilities that a particular value x belongs to either male or female and that the set of wage determinants, (X, ϵ) differ across gender. Interested readers are referred to e.g. Fortin et al. (2011) for detailed explanations of these two assumptions. $p(x)$ is the selection probability (score) for each indidual. It is estimated by a Logit model of a set of commonly employed characteristics X; see below for descriptions. Once we identify the counterfactual distributions of interest, we can then perform our counterfactual analysis using the approaches discussed above.

3. Data

To perform our analysis, we use data from the 1976-2011 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 counterfactual analysis. It thus has been widely used in the literature to study the gender gap (e.g. Waldfogel and Mayer, 2000). We begin at 1976 since it was the first year that information on weeks worked and hours worked are available in the March CPS. We restrict our sample to individuals aged between 18 and 64 who work only for wages and salary. 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 exclude part-time workers who worked less than 35 hours per week in the previous year.

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 differences in the distributions of log hourly wages between men and women are our measures of the gender gap. The differences in a specific part of the distribution can be interpreted as percentage differences. Note, however, that our distributional measure of the gender gap and SD tests are invariant to increasing monotonic transformation, while conventional measures of the gender gap are.

In our counterfactual analysis, we include age, age squared, education (four education groups: Below high school, High School, 1-3 years of College, and College and Above), current marital status (1 if non-married and zero otherwise), race (1 if non-white and zero otherwise), and region (northeast, midwest, south, and west). We also include occupations which are divided into three categories: high-skill (managerial and professional specialty occupations); medium-skill (technical, sales, and administrative support occupations); and low-skill(other occupations such as helpers, construction, and extractive occupations).

4. Results

4.1. Baseline Analysis

4.1.1. Trend of the Gender Gap 1976 - 2011

Table (1) reports a number of popular measures of the gender gap. Column (1) displays our distributional measure of the gender gap S_ρ. Recall that, S_ρ is normalized, taking values in $[0, 1]$, and to facilitate the presentation, the results reported are the original values $\times 100$ throughout the paper. The critical values based on 299 replications are reported in Table (**??**), columns (1)-(3). 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, 10th, 25th, 50th, 75th and 90th) that are commonly used in the literature.

We note that all measures imply that there exist substantial earnings differentials between men and women, and over the years. In particular, S_ρ is statistically significantly different from zero (it is larger than the critical values calculated at 99th percentiles of the bootstrapped distribution of S_ρ in all cases).

It is important to note a crucial 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. We note that our entropy measure, being a function of many moments of the earnings distributions, is able to account for increasing earnings that are accompanied by greater dispersion (inequality increasing). Indeed, our entropy measure is based on a generalized version of Theil's inequality measures.

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 do. 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 measure at the 10th percentile indicates the gap is about 37 percentage points, while the measure at the 90th percentile implies that it is more than 50 percentage points. The difference is as large as 13 percentage points. The differences at other parts of the earnings distribution indicate the gender gap is between 45 and 47 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. 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; see our example/case 3 above.

There is a further difficulty with assessment of the conventional measures when one examines the long-run trend in the gender gap. Looking at the trend from 1976 to 2011, we see a decrease in the difference between the earnings distributions of men and women over the past four decades, regardless of which measure is used. However, the decrease is not monotonic over time, and the timing of temporal deviations from the long-run trend dramatically varies across different measures used. To ease the presentation, we report the patterns of changes in different measures in Table (**??**). 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 the measure 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, 1984, 1988, 1990, 1997, and 2004). For example, the gender gap at the median increased in 1977, while the gender gap at other selected parts of the earnings distributions between men and women decreased. As a result, it is not clear why any of the conventional measures are representative of the rest of the earnings distribution and informative of the general trend of the gap gap in the society.

On the other hand, our entropy measure of the gender gap, S_ρ, provides a more complete picture of the trend. An advantage of our measure is that it does not place all weights on a particular part of the distribution, and the weights vary over time. For example, S_ρ indicates the broad gender gap decreased in 1977, which is consistent with the decrease at all parts but the 90th percentile; S_ρ indicates the gender gap increased in 1999, which is consistent with the increase in at the 10th and 75th percentiles. Our measure becomes particularly useful when the commonly used measures disagree with each other, especially during economic downturns. For example, in 2009 (the current 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 broad measure suggests that the gender gap indeed increased statistically (possibly due to worsened economic conditions), in agreement with the mean. One interesting finding is worth noting is that the gender gap at the 10th percentile fluctuates more around the trend over time, compared to the other parts of the distribution; the gender gap at the 90th percentile, although fluctuating sometimes, exhibits consistently a declining trend.

The magnitudes of S_ρ and other measures reported in Table (1) are not directly comparable. To further ease the comparisons of the patterns of the time trend implied by different measures, we normalize these measures. In particular, we first set 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 (**??**). As we can see, while both the measures at the 10th and 25th percentiles traced out the path of S_ρ in the first few years, none of the patterns implied by the conventional measures are consistent with the one by S_ρ. Our measure of the gender gap indicates a continuing trend of decline in the gender gap over time. Although this result is broadly consistent with the literature (e.g. Blau and Kahn, 2006), the rate of decline implied by our measure is much larger than that by the conventional measures. All traditional measures appear to severely *underestimate* the decline in the gender gap over time. In particular, our S_ρ implies the gender gap narrowed at average annual rate of about 11% during the period 1976-2011, while the annual rate implied by the gender difference at the median is only 5.2% during the same period, the largest among all conventional measures. Intuitively, this makes sense. If the gap at every part of the earnings distributions between men and women decreases, the decrease in the distance between two distributions should be even larger.

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). 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.

Our measure of the gender gap also uncovers another interesting finding that is masked by conventional measures: the gender gap dropped precipitously before 1990s, but the convergence drastically slowed down since 1990s. Table (2) quantifies these patterns. Specifically, the rate of decline was 7.6% before 1994, while it was only 2.5% afterwards. 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 reduce the gender gap. 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.

4.1.2. Stochastic Dominance Test Results

As discussed above, these measures of the gender gap do not lend themselves to ranking of the earnings distributions between men and women. Therefore, we now to turn to SD tests. SD results are reported in Table (3) and the corresponding comparisons of CDFs over time plotted in Figures (??)-(??). Note that the column labeled *Observed Ranking* 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.

We first notice that the earnings distribution among men lie predominantly to the right of the earnings distribution among women, indicating higher level of earnings for men. This casual observation is consistent with the fact that the differences in selected percentiles of the earnings distributions between men and women are uniformly positive. Moreover, we also notice that 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.

Our SD test statics confirm these casual observations. As we can see from Table (3), we do observe stochastic dominance throughout the whole period. In particular, the earnings distribution among men is generally found to empirically dominate, in a first-order sense, the earnings distribution among women in most cases. This result is extremely powerful: any individuals with a social welfare function in the class U_1 (as long as it is increasing in earnings) would prefer the male distribution to the female distribution, concluding that men perform better than women in the labor market. In 2002 and 2010, we find only second-order dominance relations. In other words, any individuals with a social welfare function in the class U_2 (increasing *and* concave in earnings) would conclude that men perform better than women in the labor market. However, since we do not find FSD, individuals possessing different welfare functions in the class U_1 would disagree about this conclusion. Our SD analysis makes explicit that such a ranking is possible only by accounting for "dispersion" in the welfare criteria.

That we consistently find dominance even for the beginning of the recent recessionary period (2007-2008) is surprising. As we know, industries such as construction and manufacturing where men are primary workforce are hit harder in the current recession, and as a result, men share a larger burden of the recession than women do (Sahin et al., 2010). Compared the previous recessions, we are therefore less likely to find any dominance relations in the current one.[3] However, our result indicates that although disproportionately affected, men still perform relatively better compared to women in the labor market.

Another finding is noteworthy. Despite the fact that the rate of decline has slowed down since mid-1990s, it is less likely to find statistically significant first-order dominance relations during this period. In particular, there are 6 such cases out of 18. In one case (1998), we do not even find any statistically significant dominance at all (even though we observe FSD, the p-value is only 0.72; the p-value for SSD is 0.73). The inability to rank order the earnings distributions between men and women in this case is equally informative. This finding implies that for 1998, any welfare conclusions concerning that women fare worse than men in the labor market are not robust to changes in the particular welfare function being used, despite the fact that the differences in selected percentiles of the earnings distributions between men and women are, in most cases, positive and in favor of men. This result is in stark contrast with the common belief based on the conventional measures above, illustrating the benefit to considering the *entire* distribution within the welfare economics framework when studying the gender gap.

In sum, while we consistently find dominance relations in most cases, it is less likely to conclude that men perform better than women in the labor market without explicitly taking into account "dispersion" after 1994. These results altogether suggest that women's labor market situation has improved relative to men's. Nevertheless, strong evidence of dominance relations in most cases indicates that the improvement is still far from satisfactory; this result casts doubt on the effectiveness of the welfare reforms, or any similar reforms, intended

[3]There is evidence that when the economy is further into the recession, men were likely to find jobs faster than women (partly because they work in fields previously dominated by women. Source: http://www.pewsocialtrends.org/2011/07/06/two-years-of-economic-recovery-women-lose-jobs-men-find-them/. The dominance relations are thus more likely to be found later.

to improve women's labor market outcomes. This result is consistent with the previous literature such as Blau and Kahn (2006).

4.2. Counterfactual Analysis

Our measure of the gender gap and SD results above strongly support that the gender gap remains and is in favor of men. Such conclusion is generally robust to a wide class of social welfare functions. The question then is: Can we improve women's labor market situation? Given the possibility that any policy proposed may result in both "winners" and "losers", could we still make robust statements that women would benefit from certain type of policies? Having these questions in mind, we now turn to our counterfactual analysis.

We first consider a thought experiment that would change women's pay structure. This type of thought experiment is closely related to policies aimed at promoting equal wage-setting process/structure across gender, for example, pay equity programs that are designed to addressing wage differences between men and women with the same skills and work by equalizing their pay structures. Due to data limitations, we only have a handful of human capital characteristics available in our data, as in the literature. However, we want to emphasize that such exercise is illustrative, and believe that it could still give us useful information regarding the effectiveness of these potential programs. Table (4) reports various measures of the differences between the female wage distribution and the counterfactual wage distribution (#1) corresponding to our thought experiment (i.e. the distribution of women's wages when their human capital characteristics are paid under men's wage structure). We find that the distance between two distributions is large and statistically significant. Recall that the difference captures only the difference in wage structures between men and women, while holding women's human capital characteristics constant. This result appears to be consistent with the common finding of the importance of wage structure in explaining the wage difference between men and women. The importance of structural difference, however, decreased over time. The implied annualized rate of decline is about 4 percent. In this case, we again find that the implied distance (or importance) varies across various conventional measures, and these measures severely underestimate the rate of the decline.

Turning to the SD results (Table (5)), we find that the female wage distribution and the counterfactual wage distribution # 1 are generally rankable. In most cases, we find first-order dominance relations. This result implies that, in these cases, changing earnings structure would result in a change in the earnings distribution for women, and that the change is *uniformly* in favor of all women for any individuals with social welfare functions increasing in earnings. In these cases, *all* women are winners! Such results are qualitatively consistent with the prior findings that such policies as equity pay could be potentially successful in closing the gender earnings gap (e.g. Hartmann and Aaronson, 1994; Gunderson and Riddell, 1992). However, our results are much stronger. Recall that the existing literature that typically relies on regression analysis based on means or averages ignores the rest of the distribution. Thus, it is a rather surprising yet strong result that *all* women would benefit from a policy that change their wage structure to men's.

However, we also find several cases where such result fails to hold. For example, in two cases (2002 and 2010), we fail to find FSD but do find SSD; this result means that even though there are losers and winners from this type of policy – the losers are mostly concentrated in the extreme upper tail (highest earners). As a result, any individuals with social welfare function increasing in wage and averse to inequality would still conclude there exists a welfare improvement for women from changing the wage structure; however, such conclusion may not necessarily hold for those with other social welfare functions in U_1. Furthermore, in 1986 and 1999, the dominance relations are not statistically significant. In other words, we fail to find that the female wage distribution and the counterfactual wage distribution # 1 are rankable, in sharp contrast to the conclusion based on the signs of conventional measures. This result implies that, although changing earnings structure would result in a change in the earnings distribution for women, the change is *not* uniformly in favor of all women. In the presence of both "winners" and "losers", conclusions based on previous regression analysis may not be robust at all. We emphasize this result: For any individuals with social welfare functions either in U_1 (increasing in wage) or U_2 (increasing in wage and averse to inequality), the change resulted from this type of policy does not necessarily represent a welfare improvement for women. Failure to find any statistically meaningful SD implies that the conclusion of welfare improvement cannot be obtained without further restriction on welfare functions.

We now consider another thought experiment resembling policies aimed at changing women's human capital characteristics such as education. In this thought experiment, we shall contrast the female wage distribution with the counterfactual wage distribution (#2) (i.e. the distribution of women's wages when they possess men's human capital characteristics but holding women's wage structure unchanged). Table (6) reports various measures of the differences between these two distributions. In sharp contrast to the structural difference above, we find that the compositional difference – difference between the female wage distribution and the counterfactual wage distribution (#2) – is, albeit still statistically significant, rather small. In most cases, the magnitude of the compositional difference is only about one tenth of that of the structural effect. However, the magnitude has been increasing over time. The annualized rate of increase is also about 4 percent. Moreover, unlike in the case of structural difference above, we find that conventional measures trace out the pattern of our distributional measure well.

Turning to the SD results (Table (7)), we fail to find statistically meaningful SD ranking of the female wage distribution and the counterfactual wage distribution (#2) in more than half of the period (14 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 from the societal point of view. This result is also consistent with the above finding of a rather small difference between these two distributions. Further, in some cases where we do find dominance relations, the female wage distribution actually dominates the counterfactual distribution. It implies that women could be even worse off when they have the same distribution of human capital characteristics as do men. Altogether, our results imply 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.

5. Conclusions

In this paper, we present a set of complementary tools that move beyond the simple moment-based comparison of the earnings distributions. In particular, we propose a new measure of the gender gap based on the the distance between two *whole* earnings distributions, instead of their specific parts. We also introduce tests based on stochastic dominance to allow for robust welfare comparisons of the earnings distributions between men and women. Using the CPS data 1976-2011, we apply this framework to analyze the gender gap in the U.S. We reach several main conclusions. First, we find that traditional moment-based measures severely underestimate the declining trend of the gender gap in the U.S. Second, we find that even though first-order stochastic dominance exists in most cases, there are a few cases where we find only second-order dominance or even no dominance; this result highlights the importance in taking into account welfare criteria to evaluate the gender gap situation. Finally, we further compare the female wage distribution to two counterfactual distributions (the earnings that women would earn under the earnings structure of men and the earnings that women would earn should they have men's characteristics). We find that structural effects are generally more important than composition effects. However, the importance of structural effects has declined over time, while that of composition effects has increased. Our SD results indicate that policies aimed at changing women's pay structure are more likely to improve women's earnings prospects than policies aimed at changing women's human capital characteristics. Even though this conclusion is largely true, we do find several cases where in the presence of "winners" and "losers", we cannot draw definite conclusions regarding the effectiveness of such policies unless we are willing to impose stronger and more restrictive assumptions on social welfare functions. These results showcase the importance of distributional aspects in measuring the gender gap, as well as taking into welfare criteria in defining and testing rankings between two distributions.

References

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

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.

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

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.

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

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

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.

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.

Li, Q. and J. Racine. 2007. *Nonparametric Econometrics*. Princeton, New Jersey: Princeton University Press.

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

Lugo, M.A. and E. Maasoumi. 2008. "Multidimensional Poverty Measures from an Information Theory Perspective." *ECINEQ Working Paper* 2008.

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.

Mas-Colell, A., M.D. Whinston, and J.R. Green. 1995. *Microeconomic Theory*. Oxford University Press.

Millimet, D.L. and L. Wang. 2006. "A Distributional Analysis of the Gender Earnings Gap in Urban China." *The B.E. Journal of Economic Analysis & Policy (Contributions)* 5:Article 5.

Polachek, S.W. 2006. *The Declining Significance of Gender?*, chapter How the Life-Cycle Human-Capital Model Explains Why The Gender Wage Gap Narrowed, pp. 102–124. New York: Russell Sage Foundation.

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

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. 36377. New York: Springer.

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

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* .

Table 1: Measures of The Gender Gaps

Year	$S_\rho \times 100$ (1)	Mean (2)	10th (3)	25th (4)	50th (5)	75th (6)	90th (7)
1976	10.978	0.466	0.362	0.45	0.474	0.472	0.505
1977	10.252	0.444	0.318	0.414	0.488	0.470	0.480
1978	10.155	0.446	0.325	0.416	0.455	0.483	0.474
1979	9.891	0.436	0.288	0.386	0.473	0.495	0.484
1980	9.469	0.422	0.266	0.366	0.472	0.485	0.455
1981	8.957	0.403	0.241	0.354	0.442	0.474	0.459
1982	8.682	0.409	0.270	0.370	0.460	0.475	0.453
1983	7.576	0.387	0.229	0.298	0.433	0.431	0.464
1984	6.505	0.354	0.193	0.297	0.381	0.409	0.405
1985	6.190	0.355	0.209	0.29	0.392	0.414	0.419
1986	5.612	0.339	0.213	0.30	0.373	0.405	0.393
1987	5.009	0.333	0.231	0.297	0.355	0.389	0.382
1988	4.606	0.316	0.199	0.265	0.344	0.375	0.344
1989	4.425	0.315	0.223	0.292	0.359	0.354	0.350
1990	3.602	0.291	0.204	0.241	0.311	0.325	0.317
1991	3.134	0.269	0.163	0.244	0.288	0.306	0.316
1992	2.859	0.259	0.148	0.222	0.285	0.288	0.316
1993	2.662	0.241	0.136	0.21	0.235	0.295	0.289
1994	2.284	0.238	0.146	0.197	0.254	0.276	0.273
1995	2.329	0.248	0.173	0.206	0.260	0.288	0.272
1996	2.259	0.248	0.175	0.218	0.250	0.274	0.272
1997	2.173	0.238	0.160	0.191	0.248	0.262	0.266
1998	2.206	0.243	0.145	0.236	0.258	0.251	0.280
1999	2.320	0.242	0.149	0.230	0.232	0.268	0.274
2000	2.019	0.242	0.158	0.210	0.236	0.273	0.262
2001	2.384	0.266	0.214	0.223	0.266	0.301	0.313
2002	2.271	0.256	0.176	0.239	0.243	0.253	0.310
2003	2.139	0.245	0.182	0.191	0.239	0.262	0.291
2004	1.846	0.234	0.178	0.190	0.218	0.259	0.283
2005	1.908	0.234	0.168	0.182	0.216	0.255	0.288
2006	1.813	0.234	0.173	0.208	0.210	0.254	0.277
2007	1.588	0.224	0.163	0.195	0.244	0.254	0.268
2008	1.582	0.212	0.173	0.167	0.223	0.236	0.258
2009	1.640	0.215	0.154	0.172	0.193	0.222	0.273
2010	1.566	0.211	0.157	0.185	0.203	0.235	0.269
2011	1.436	0.202	0.172	0.185	0.182	0.259	0.248

[1] Data Source: IPUMS CPS (`http://cps.ipums.org/cps/`). Column (1) reports the overal gender gap ($\times 100$) at corresponding functionals of the distributions of log wages (measures the distance between the female and male wage distributions). Columns (2)-(6) report conventional measures based on difference in parts of the wage distributions between males and females.

Table 2: Implied Rate of Decline by Measures of The Gender Gaps

Year	$S_\rho \times 100$ (1)	Mean (2)	10th (3)	25th (4)	50th (5)	75th (6)	90th (7)
Pre-1994	-0.076	-0.036	-0.053	-0.041	-0.038	-0.026	-0.031
Post-1994	-0.025	-0.009	0.009	-0.003	-0.018	-0.004	-0.005
Whole Period	-0.107	-0.045	-0.040	-0.048	-0.052	-0.033	-0.039

[1] The calculations are based on Table (1).

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

Year	Observed Ranking	$d_{1,max}$	$d_{2,max}$	d	$Pr[d \leq 0]$	$s_{1,max}$	$s_{2,max}$	s	$Pr[s \leq 0]$
1976	FSD	54.68	-0.15	-0.15	1.00	2858.89	-0.15	-0.15	1.00
1977	FSD	57.97	-0.16	-0.16	1.00	2898.34	-0.16	-0.16	1.00
1978	FSD	57.07	-0.16	-0.16	1.00	2926.32	-0.16	-0.16	1.00
1979	FSD	57.77	-0.16	-0.16	1.00	2866.95	-0.16	-0.16	1.00
1980	FSD	61.66	-0.16	-0.16	1.00	2965.85	-0.17	-0.17	1.00
1981	FSD	59.44	-0.17	-0.17	0.99	2785.31	-0.17	-0.17	0.99
1982	FSD	54.43	-0.18	-0.18	1.00	2747.44	-0.18	-0.18	1.00
1983	FSD	49.74	-0.16	-0.16	1.00	2540.40	-0.17	-0.17	1.00
1984	FSD	46.47	-0.17	-0.17	0.99	2190.37	-0.21	-0.21	1.00
1985	FSD	46.10	-0.17	-0.17	1.00	2273.42	-0.17	-0.17	1.00
1986	FSD	43.40	-0.15	-0.15	0.99	2059.96	-0.17	-0.17	0.99
1987	FSD	41.47	-0.15	-0.15	1.00	2039.91	-0.16	-0.16	1.00
1988	FSD	40.25	-0.17	-0.17	1.00	2048.63	-0.17	-0.17	1.00
1989	FSD	37.94	-0.14	-0.14	1.00	1981.06	-0.16	-0.16	1.00
1990	FSD	35.94	-0.15	-0.15	1.00	3045.07	-0.17	-0.17	1.00
1991	FSD	33.32	-0.15	-0.15	1.00	2576.40	-0.19	-0.19	1.00
1992	FSD	31.99	-0.16	-0.16	1.00	1634.84	-0.16	-0.16	1.00
1993	FSD	30.24	-0.17	-0.17	0.97	1585.04	-0.19	-0.19	0.97
1994	FSD	27.94	-0.16	-0.16	1.00	1739.99	-0.17	-0.17	1.00
1995	FSD	28.74	-0.14	-0.14	1.00	1851.10	-0.16	-0.16	1.00
1996	FSD	25.73	-0.15	-0.15	1.00	1584.94	-0.15	-0.15	1.00
1997	FSD	26.14	-0.17	-0.17	0.93	1292.97	-0.17	-0.17	0.93
1998	FSD	26.18	-0.10	-0.10	0.72	1403.89	-0.16	-0.16	0.73
1999	FSD	26.50	-0.12	-0.12	0.57	1542.11	-0.17	-0.17	0.97
2000	FSD	25.93	-0.16	-0.16	1.00	1596.50	-0.16	-0.16	1.00
2001	FSD	34.60	-0.21	-0.21	1.00	2010.59	-0.21	-0.21	1.00
2002	SSD	33.76	0.29	0.29	0.03	1908.06	-0.20	-0.20	1.00
2003	FSD	31.19	-0.20	-0.20	0.88	1700.05	-0.20	-0.20	1.00
2004	FSD	30.20	-0.17	-0.17	1.00	1865.61	-0.21	-0.21	1.00
2005	FSD	30.10	-0.18	-0.18	1.00	1714.16	-0.20	-0.20	1.00
2006	FSD	29.25	-0.20	-0.20	1.00	1570.70	-0.21	-0.21	1.00
2007	FSD	27.94	-0.18	-0.18	1.00	1634.74	-0.20	-0.20	1.00
2008	FSD	27.49	-0.26	-0.26	0.99	1863.03	-0.26	-0.26	1.00
2009	FSD	27.64	-0.20	-0.20	0.88	1528.66	-0.20	-0.20	0.93
2010	SSD	27.29	0.30	0.30	0.02	1471.40	-0.19	-0.19	1.00
2011	FSD	25.04	-0.17	-0.17	0.95	1407.96	-0.19	-0.19	1.00

Table 4: Measures of Differences between Female and Female counterfactual #1 Distributions

Year	$S_\rho \times 100$ (1)	Mean (2)	10th (3)	25th (4)	50th (5)	75th (6)	90th (7)
1976	8.998	0.417	0.317	0.383	0.431	0.438	0.460
1977	8.517	0.399	0.268	0.376	0.427	0.427	0.431
1978	8.565	0.403	0.276	0.349	0.412	0.448	0.452
1979	8.438	0.400	0.248	0.347	0.433	0.454	0.470
1980	8.400	0.402	0.262	0.340	0.432	0.468	0.436
1981	8.117	0.383	0.223	0.332	0.417	0.452	0.436
1982	7.885	0.389	0.270	0.341	0.428	0.438	0.428
1983	6.709	0.363	0.221	0.293	0.405	0.400	0.423
1984	6.151	0.346	0.199	0.293	0.375	0.397	0.392
1985	6.052	0.355	0.214	0.290	0.392	0.394	0.419
1986	5.457	0.334	0.218	0.300	0.373	0.405	0.393
1987	4.853	0.333	0.261	0.297	0.346	0.368	0.378
1988	4.636	0.323	0.228	0.274	0.344	0.375	0.340
1989	4.791	0.334	0.239	0.305	0.363	0.358	0.368
1990	4.181	0.319	0.257	0.288	0.336	0.340	0.339
1991	3.782	0.302	0.208	0.277	0.327	0.320	0.334
1992	3.622	0.299	0.198	0.282	0.310	0.315	0.316
1993	3.401	0.283	0.195	0.257	0.289	0.331	0.325
1994	3.036	0.280	0.223	0.257	0.300	0.312	0.307
1995	2.922	0.281	0.191	0.251	0.303	0.288	0.305
1996	3.050	0.294	0.210	0.272	0.297	0.304	0.316
1997	3.059	0.294	0.240	0.271	0.297	0.297	0.310
1998	3.101	0.296	0.210	0.288	0.292	0.307	0.344
1999	3.388	0.298	0.244	0.279	0.297	0.308	0.313
2000	3.165	0.308	0.248	0.287	0.307	0.332	0.318
2001	3.176	0.310	0.243	0.275	0.320	0.309	0.366
2002	3.215	0.309	0.240	0.288	0.288	0.313	0.364
2003	2.862	0.287	0.234	0.234	0.267	0.312	0.342
2004	2.641	0.283	0.227	0.250	0.273	0.312	0.312
2005	2.900	0.295	0.232	0.223	0.288	0.319	0.348
2006	2.811	0.296	0.223	0.262	0.288	0.323	0.359
2007	2.531	0.283	0.220	0.260	0.283	0.305	0.331
2008	2.640	0.279	0.242	0.243	0.279	0.300	0.324
2009	2.680	0.281	0.248	0.245	0.278	0.296	0.357
2010	2.589	0.278	0.223	0.262	0.280	0.307	0.325
2011	2.119	0.249	0.223	0.237	0.239	0.304	0.297

[1] Data Source: IPUMS CPS (`http://cps.ipums.org/cps/`). Column (1) reports the overal gender gap ($\times 100$) at corresponding functionals of the distributions of log wages (measures the distance between the female and female counterfactual #1 wage distributions). Columns (2)- (6) report conventional measures based on difference in parts of between the female and female counterfactual #1 wage distributions. Female counterfactual #1 distribution is the counterfactual wage distribution among women when their human characteristics are valued under men's wage structure.

Table 5: Stochastic Dominance Results (Female v.s. Female Counterfactual #1 Wage Distributions)

Year	Observed Ranking	$d_{1,max}$	$d_{2,max}$	d	$Pr[d \leq 0]$	$s_{1,max}$	$s_{2,max}$	s	$Pr[s \leq 0]$
1976	FSD	48.80	-0.14	-0.14	1.00	2562.04	-0.14	-0.14	1.00
1977	FSD	52.75	-0.16	-0.16	1.00	2605.25	-0.16	-0.16	1.00
1978	FSD	52.08	-0.15	-0.15	0.96	2646.30	-0.16	-0.16	0.96
1979	FSD	53.30	-0.11	-0.11	1.00	2629.37	-0.17	-0.17	1.00
1980	FSD	58.25	-0.16	-0.16	1.00	2817.56	-0.17	-0.17	1.00
1981	FSD	56.46	-0.17	-0.17	1.00	2644.61	-0.18	-0.18	1.00
1982	FSD	51.51	-0.16	-0.16	0.99	2612.90	-0.16	-0.16	1.00
1983	FSD	46.69	-0.15	-0.15	1.00	2380.20	-0.17	-0.17	1.00
1984	FSD	45.11	-0.16	-0.16	0.98	2145.20	-0.17	-0.17	0.98
1985	FSD	45.91	-0.16	-0.16	1.00	2268.15	-0.16	-0.16	1.00
1986	FSD	42.81	-0.21	-0.21	0.76	2045.17	-0.21	-0.21	0.77
1987	FSD	40.69	-0.14	-0.14	0.98	2040.58	-0.18	-0.18	0.99
1988	FSD	40.49	-0.17	-0.17	1.00	2091.63	-0.17	-0.17	1.00
1989	FSD	39.33	-0.16	-0.16	1.00	2096.96	-0.18	-0.18	1.00
1990	FSD	39.34	-0.16	-0.16	1.00	3336.83	-0.17	-0.17	1.00
1991	FSD	37.25	-0.17	-0.17	1.00	2887.26	-0.18	-0.18	1.00
1992	FSD	36.24	-0.16	-0.16	1.00	1889.86	-0.17	-0.17	1.00
1993	FSD	34.45	-0.16	-0.16	0.99	1858.63	-0.22	-0.22	0.99
1994	FSD	32.44	-0.15	-0.15	0.99	2051.06	-0.16	-0.16	0.99
1995	FSD	32.31	-0.16	-0.16	1.00	2094.29	-0.17	-0.17	1.00
1996	FSD	30.16	-0.11	-0.11	1.00	1877.99	-0.15	-0.15	1.00
1997	FSD	31.34	-0.15	-0.15	1.00	1589.95	-0.16	-0.16	1.00
1998	FSD	30.86	-0.17	-0.17	0.91	1703.95	-0.17	-0.17	0.92
1999	FSD	32.35	-0.15	-0.15	0.38	1902.00	-0.17	-0.17	0.87
2000	FSD	33.06	-0.17	-0.17	1.00	2034.39	-0.17	-0.17	1.00
2001	FSD	40.36	-0.18	-0.18	1.00	2347.71	-0.20	-0.20	1.00
2002	SSD	40.70	0.28	0.28	0.07	2301.80	-0.20	-0.20	1.00
2003	FSD	36.60	-0.21	-0.21	0.97	1996.78	-0.21	-0.21	1.00
2004	FSD	36.12	-0.24	-0.24	1.00	2261.05	-0.24	-0.24	1.00
2005	FSD	37.79	-0.18	-0.18	1.00	2158.72	-0.20	-0.20	1.00
2006	FSD	37.17	-0.19	-0.19	1.00	1988.91	-0.19	-0.19	1.00
2007	FSD	35.82	-0.17	-0.17	1.00	2069.90	-0.20	-0.20	1.00
2008	FSD	36.34	-0.19	-0.19	0.99	2464.95	-0.20	-0.20	0.99
2009	FSD	35.54	-0.20	-0.20	0.95	1995.00	-0.21	-0.21	1.00
2010	SSD	35.37	0.29	0.29	0.05	1939.45	-0.20	-0.20	1.00
2011	FSD	30.61	-0.19	-0.19	0.97	1735.31	-0.19	-0.19	1.00

Table 6: Measures of Differences between Female and Female counterfactual #2 Distributions

Year	$S_\rho \times 100$ (1)	Mean (2)	10th (3)	25th (4)	50th (5)	75th (6)	90th (7)
1976	0.117	-0.026	-0.038	-0.032	-0.050	-0.023	0.000
1977	0.124	-0.030	-0.012	-0.041	-0.047	-0.010	0.000
1978	0.121	-0.032	-0.059	-0.048	-0.049	-0.017	0.000
1979	0.068	-0.022	-0.038	-0.036	-0.035	-0.018	0.000
1980	0.118	-0.041	-0.039	-0.049	-0.026	-0.025	-0.008
1981	0.155	-0.037	-0.079	-0.067	-0.028	-0.010	-0.003
1982	0.159	-0.041	-0.066	-0.051	-0.047	-0.029	-0.001
1983	0.158	-0.043	-0.065	-0.084	-0.036	-0.024	0.000
1984	0.157	-0.041	-0.057	-0.065	-0.069	-0.028	-0.025
1985	0.206	-0.057	-0.065	-0.078	-0.057	-0.043	0.000
1986	0.179	-0.052	-0.051	-0.051	-0.059	-0.004	-0.021
1987	0.204	-0.054	-0.053	-0.065	-0.069	-0.037	-0.004
1988	0.221	-0.061	-0.107	-0.095	-0.057	-0.031	-0.013
1989	0.246	-0.064	-0.102	-0.111	-0.065	-0.044	-0.028
1990	0.240	-0.069	-0.086	-0.107	-0.069	-0.046	-0.026
1991	0.334	-0.082	-0.100	-0.104	-0.116	-0.066	-0.036
1992	0.323	-0.078	-0.086	-0.109	-0.099	-0.075	-0.029
1993	0.297	-0.077	-0.077	-0.111	-0.104	-0.055	-0.031
1994	0.305	-0.084	-0.082	-0.111	-0.094	-0.068	-0.020
1995	0.242	-0.070	-0.098	-0.108	-0.079	-0.053	-0.029
1996	0.268	-0.075	-0.086	-0.088	-0.091	-0.047	-0.038
1997	0.284	-0.074	-0.073	-0.129	-0.078	-0.039	-0.046
1998	0.309	-0.084	-0.078	-0.079	-0.113	-0.086	-0.039
1999	0.369	-0.090	-0.094	-0.123	-0.125	-0.072	-0.059
2000	0.326	-0.086	-0.095	-0.078	-0.095	-0.059	-0.041
2001	0.232	-0.069	-0.061	-0.093	-0.056	-0.056	-0.039
2002	0.264	-0.078	-0.073	-0.090	-0.085	-0.072	-0.045
2003	0.204	-0.067	-0.080	-0.103	-0.097	-0.051	-0.037
2004	0.280	-0.081	-0.100	-0.097	-0.091	-0.050	-0.069
2005	0.357	-0.092	-0.119	-0.123	-0.109	-0.049	-0.049
2006	0.342	-0.091	-0.074	-0.105	-0.128	-0.082	-0.059
2007	0.340	-0.093	-0.105	-0.085	-0.084	-0.079	-0.069
2008	0.334	-0.087	-0.079	-0.095	-0.079	-0.074	-0.061
2009	0.336	-0.093	-0.077	-0.130	-0.105	-0.105	-0.074
2010	0.305	-0.088	-0.065	-0.104	-0.125	-0.086	-0.041
2011	0.487	-0.112	-0.065	-0.152	-0.154	-0.093	-0.088

[1] Data Source: IPUMS CPS (http://cps.ipums.org/cps/). Column (1) reports the overal gender gap ($\times 100$) at corresponding functionals of the distributions of log wages (measures the distance between the female and female counterfactual #2 wage distributions). Columns (2)- (6) report conventional measures based on difference in parts of between the female and female counterfactual #2 wage distributions. Female counterfactual #2 distribution is the counterfactual wage distribution among women when men's human characteristics are instead valued under women's wage structure.

Table 7: Stochastic Dominance Results (Female v.s. Female Counterfactual #2 Wage Distributions)

Year	Observed Ranking	$d_{1,max}$	$d_{2,max}$	d	$Pr[d \leq 0]$	$s_{1,max}$	$s_{2,max}$	s	$Pr[s \leq 0]$
1976	None	0.29	5.59	0.29	0.01	2.32	174.48	2.32	0.23
1977	SSD	0.47	6.50	0.47	0.02	-0.17	207.86	-0.17	0.61
1978	SSD	0.51	6.58	0.51	0.02	-0.16	214.68	-0.16	0.79
1979	SSD	0.43	4.60	0.43	0.01	-0.17	147.37	-0.17	0.41
1980	SSD	0.16	6.47	0.16	0.19	-0.17	261.51	-0.17	0.98
1981	SSD	0.06	8.14	0.06	0.11	-0.17	256.43	-0.17	0.90
1982	SSD	0.18	7.69	0.18	0.15	-0.16	276.23	-0.16	0.97
1983	FSD	0.00	7.23	0.00	0.18	-0.17	270.50	-0.17	0.98
1984	FSD	-0.11	7.42	-0.11	0.24	-0.19	265.35	-0.19	0.45
1985	FSD	-0.08	8.55	-0.08	0.40	-0.17	356.83	-0.17	0.94
1986	FSD	-0.17	8.26	-0.17	0.45	-0.17	315.12	-0.17	0.92
1987	FSD	-0.15	8.67	-0.15	0.42	-0.17	338.91	-0.17	0.89
1988	FSD	-0.16	9.23	-0.16	0.59	-0.16	390.16	-0.16	0.88
1989	FSD	-0.11	9.59	-0.11	0.31	-0.17	400.09	-0.17	0.96
1990	FSD	-0.17	9.68	-0.17	0.72	-0.17	712.75	-0.17	0.95
1991	FSD	-0.14	12.27	-0.14	0.80	-0.17	774.47	-0.17	0.97
1992	FSD	-0.11	11.75	-0.11	0.65	-0.17	494.34	-0.17	0.94
1993	FSD	-0.17	11.12	-0.17	0.73	-0.20	510.55	-0.20	0.95
1994	FSD	-0.16	10.91	-0.16	0.58	-0.17	602.71	-0.17	1.00
1995	FSD	-0.16	9.86	-0.16	0.66	-0.18	515.70	-0.18	0.89
1996	FSD	-0.10	10.21	-0.10	0.57	-0.16	463.37	-0.16	0.85
1997	FSD	-0.15	10.02	-0.15	0.42	-0.17	405.20	-0.17	0.69
1998	FSD	-0.17	11.11	-0.17	0.74	-0.20	476.77	-0.20	0.83
1999	FSD	-0.11	12.03	-0.11	0.81	-0.16	568.01	-0.16	0.91
2000	FSD	-0.16	11.54	-0.16	0.82	-0.18	569.25	-0.18	0.88
2001	FSD	-0.21	12.41	-0.21	0.91	-0.24	519.59	-0.24	0.96
2002	FSD	-0.19	12.55	-0.19	0.97	-0.20	574.93	-0.20	0.99
2003	FSD	-0.21	11.00	-0.21	0.77	-0.25	460.15	-0.25	0.86
2004	FSD	-0.20	13.07	-0.20	0.93	-0.21	636.64	-0.21	0.96
2005	FSD	-0.17	14.22	-0.17	0.93	-0.19	668.74	-0.19	0.95
2006	FSD	-0.20	14.24	-0.20	0.92	-0.20	613.36	-0.20	0.93
2007	FSD	-0.19	14.09	-0.19	0.91	-0.20	669.56	-0.20	0.94
2008	FSD	-0.21	13.70	-0.21	0.77	-0.31	770.09	-0.31	0.81
2009	FSD	-0.17	14.04	-0.17	0.96	-0.20	638.77	-0.20	0.99
2010	FSD	-0.19	13.22	-0.19	0.94	-0.19	600.92	-0.19	0.97
2011	FSD	-0.20	15.98	-0.20	0.96	-0.20	777.12	-0.20	0.98

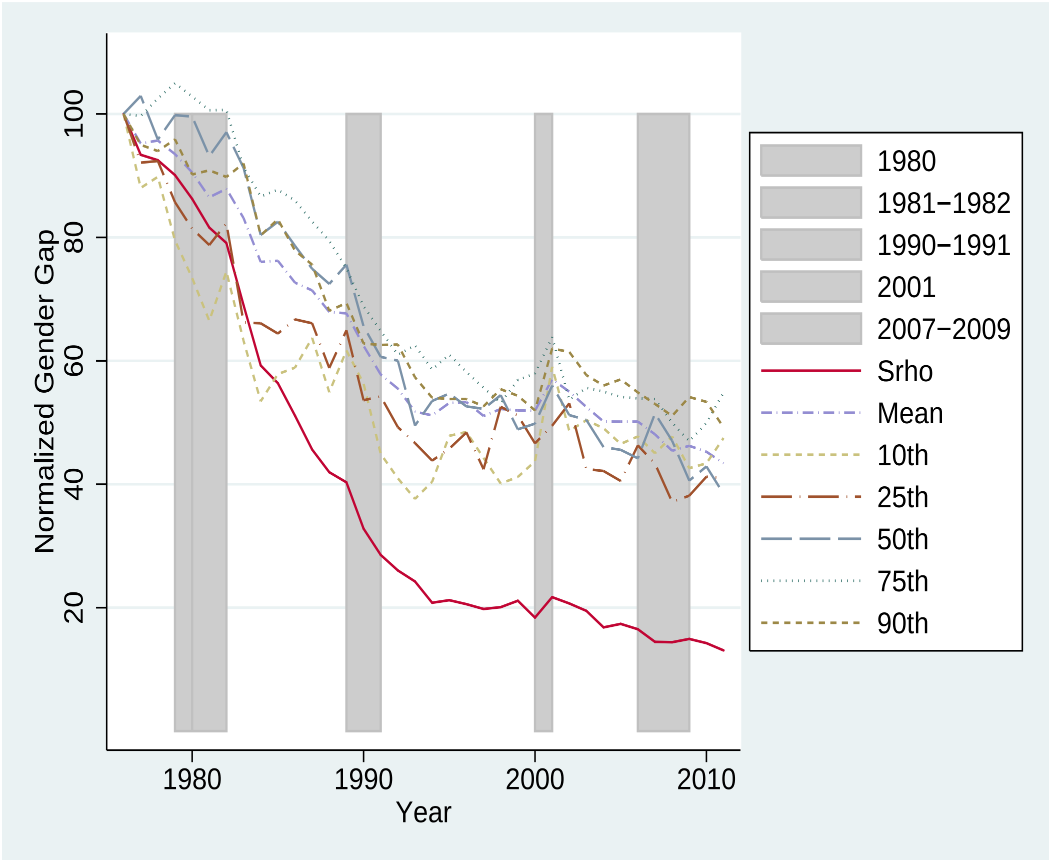
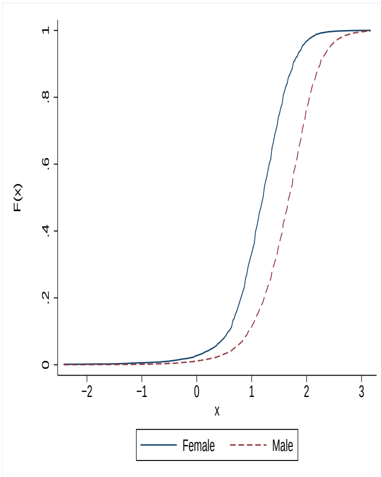
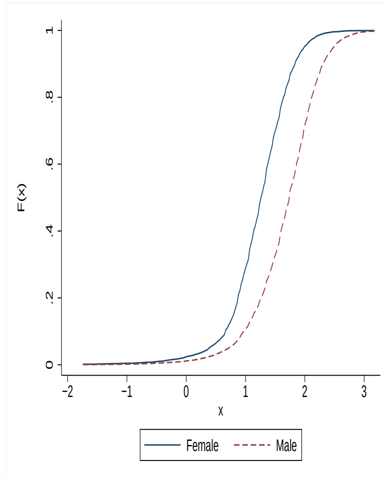


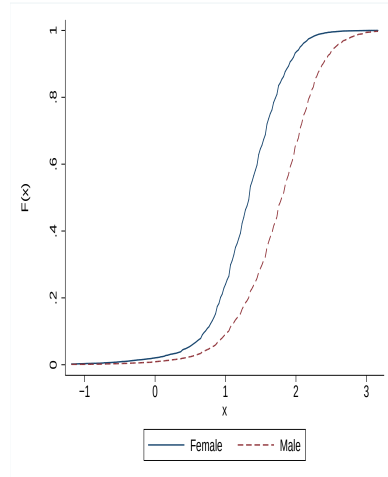
Figure 1: The Trend of Gender Wage Gap (Shaded Areas correspond to the recession periods announced by NBER)



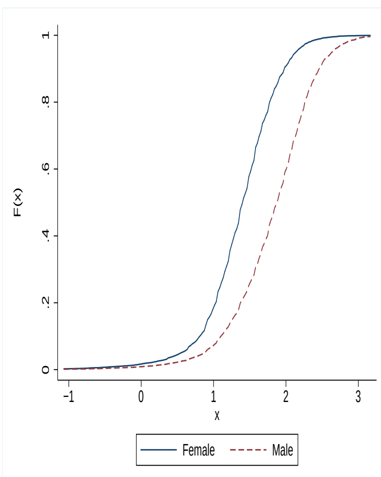
(a) 1976



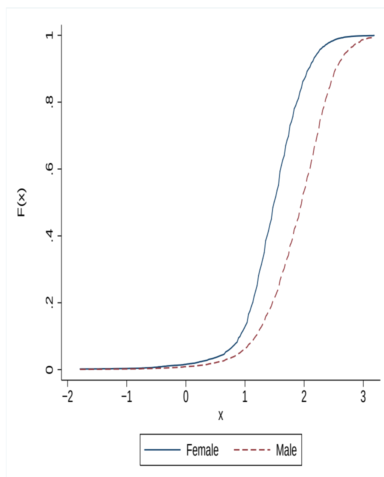
(b) 1977



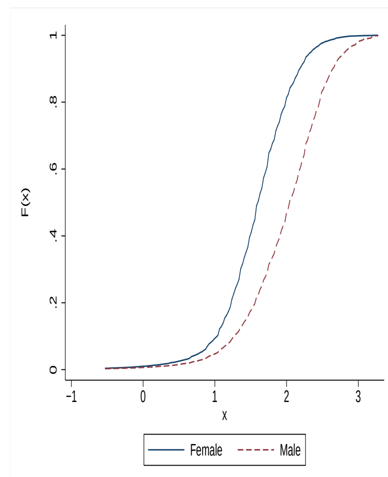
(c) 1978



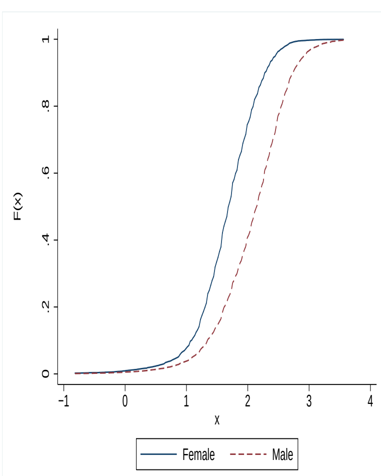
(d) 1979



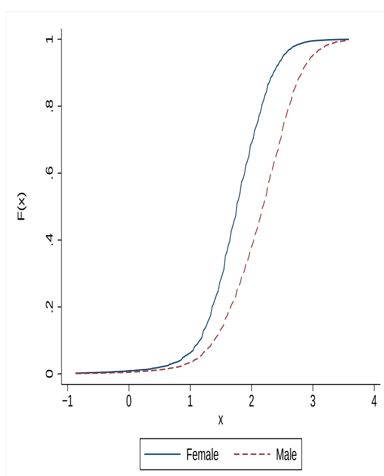
(e) 1980



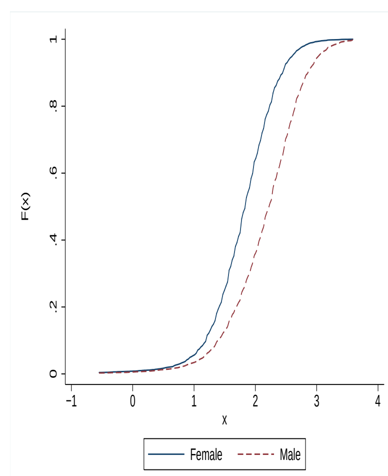
(f) 1981



(g) 1982

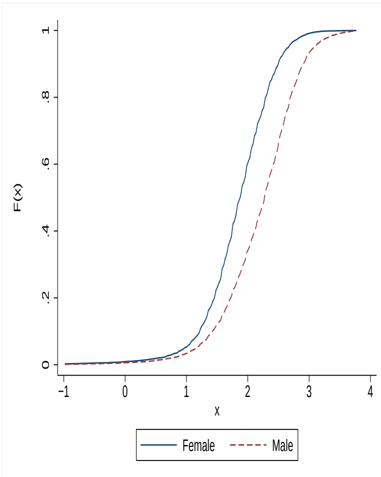


(h) 1983

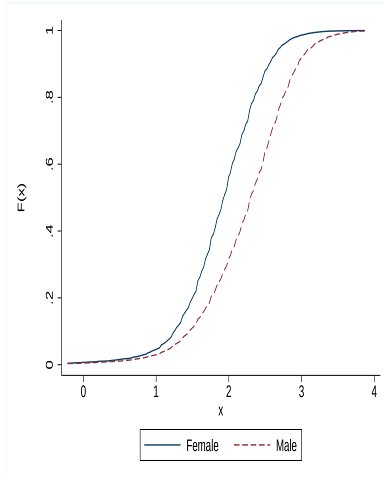


(i) 1984

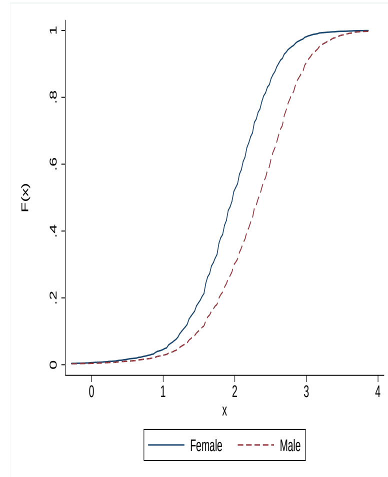
Figure 2: CDF Comparisons of Female and Male Wage Distributions (1976-1984)



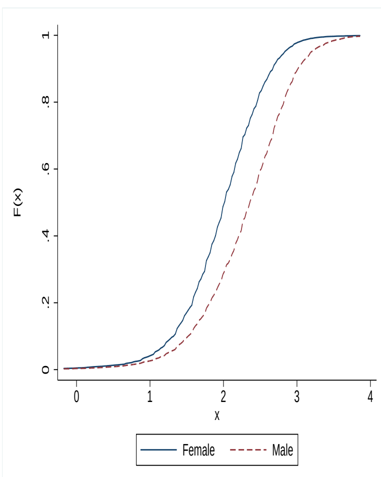
(a) 1985



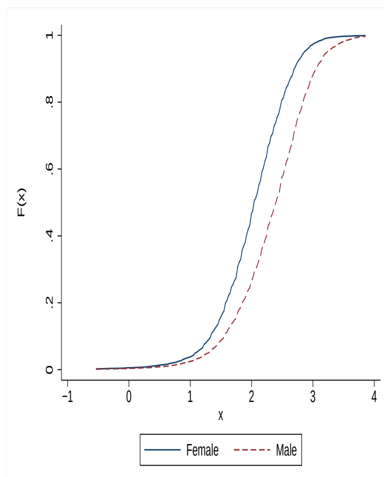
(b) 1986



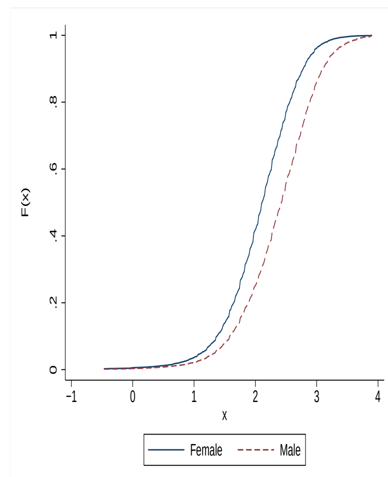
(c) 1987



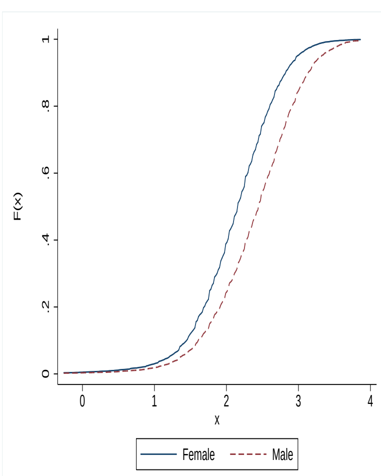
(d) 1988



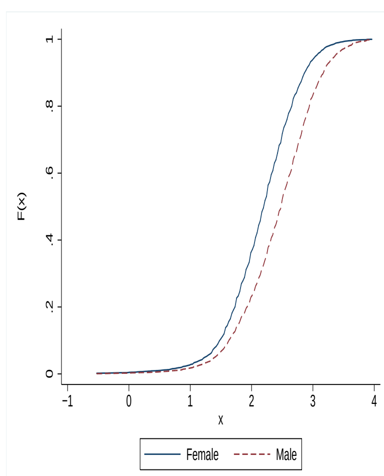
(e) 1989



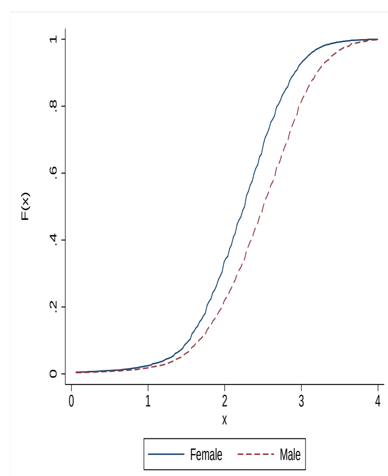
(f) 1990



(g) 1991

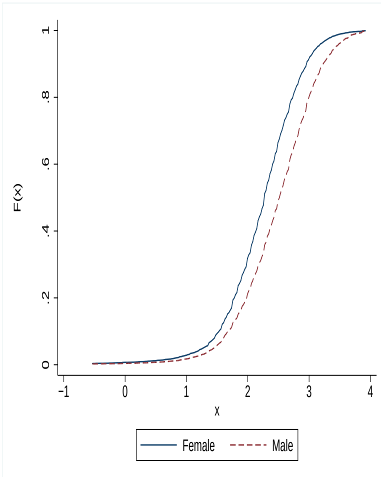


(h) 1992

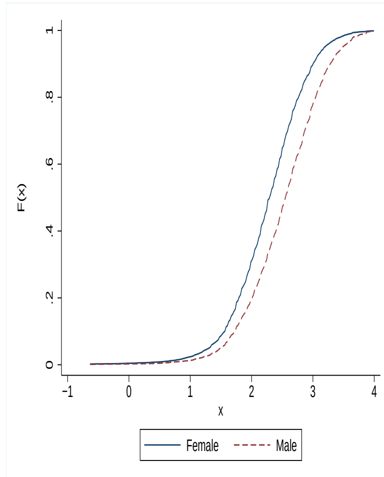


(i) 1993

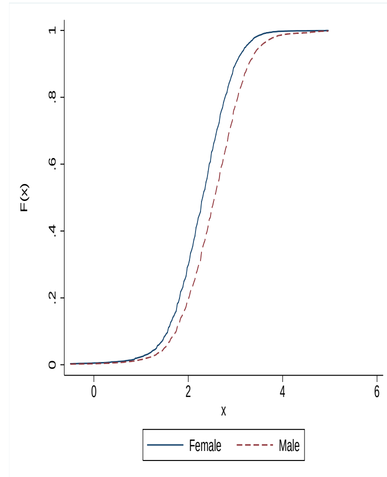
Figure 3: CDF Comparisons of Female and Male Wage Distributions (1985 - 1993)



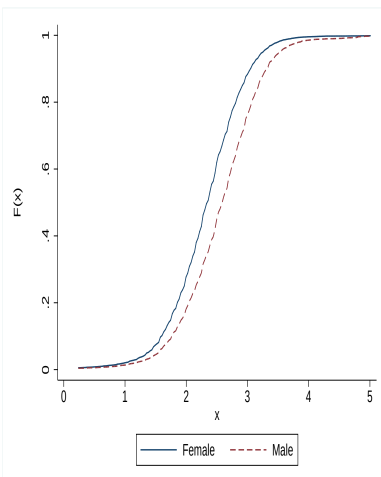
(a) 1994



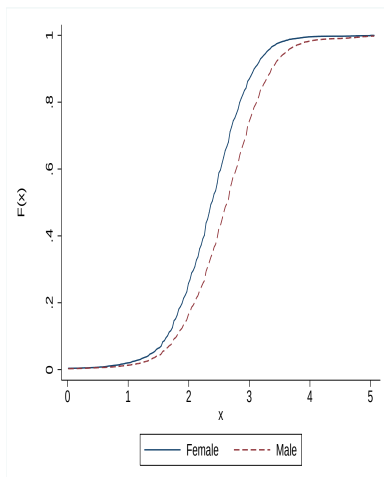
(b) 1995



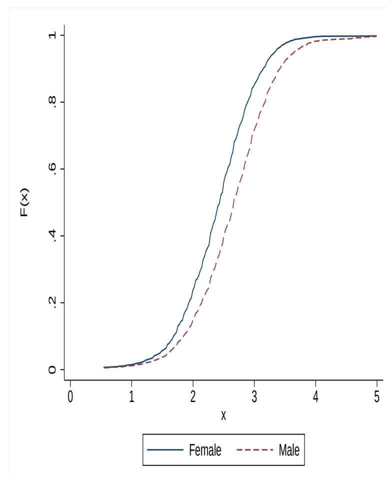
(c) 1996



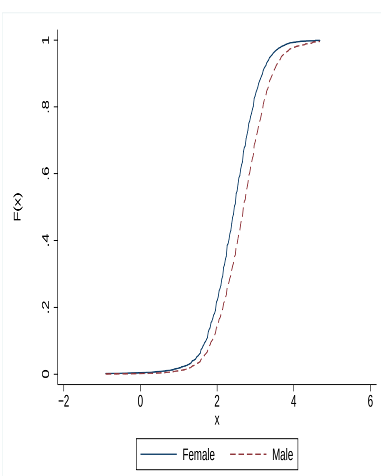
(d) 1997



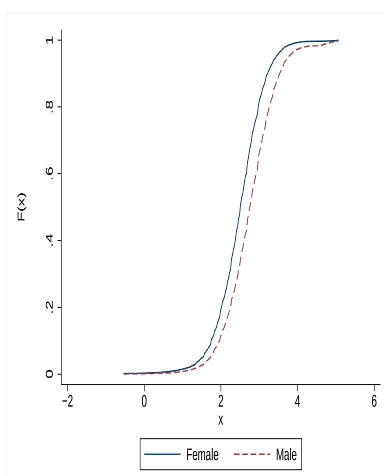
(e) 1998



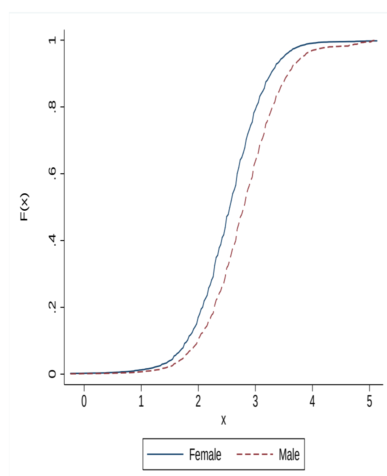
(f) 1999



(g) 2000

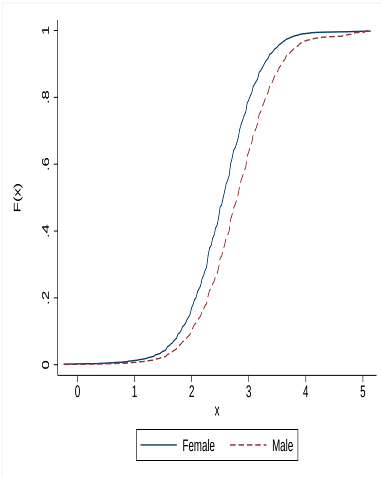


(h) 2001

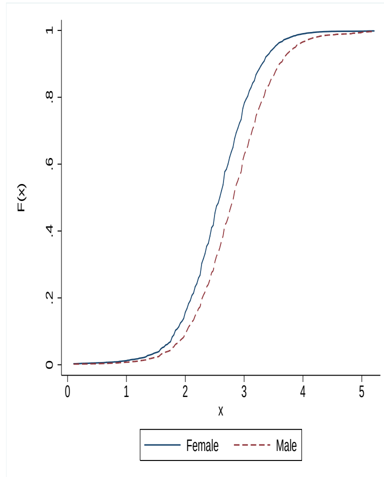


(i) 2002

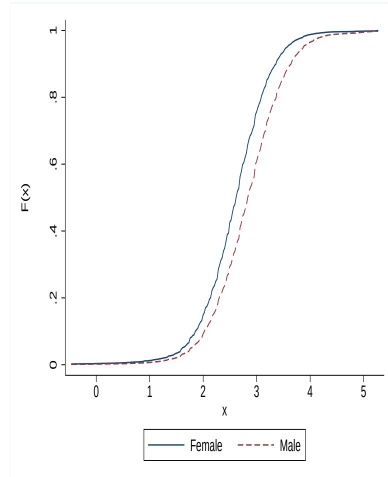
Figure 4: CDF Comparisons of Female and Male Wage Distributions (1994 - 2002)



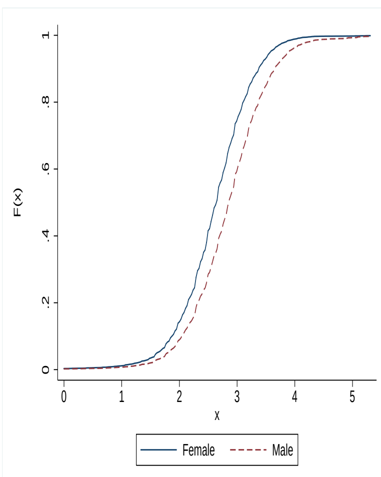
(a) 2002



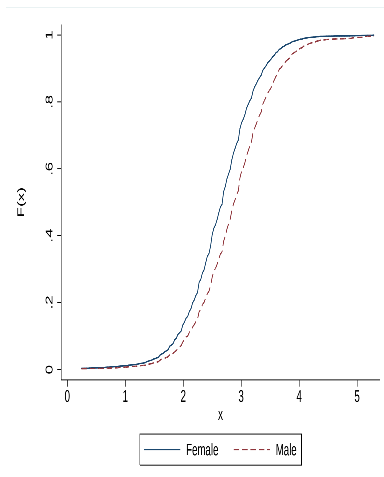
(b) 2003



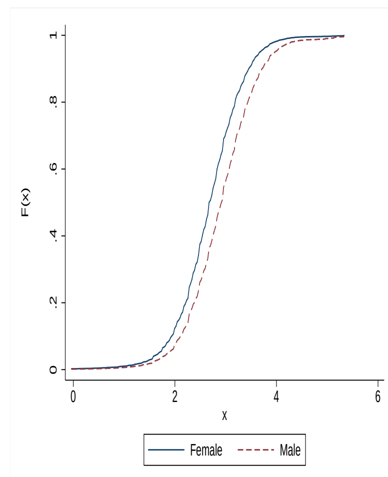
(c) 2004



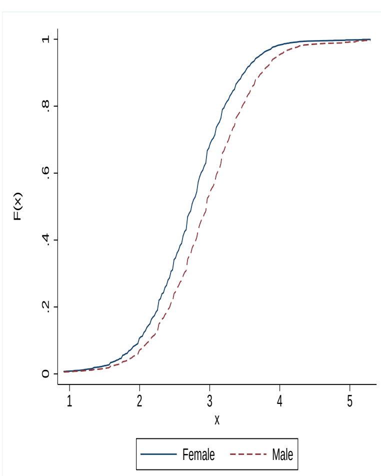
(d) 2005



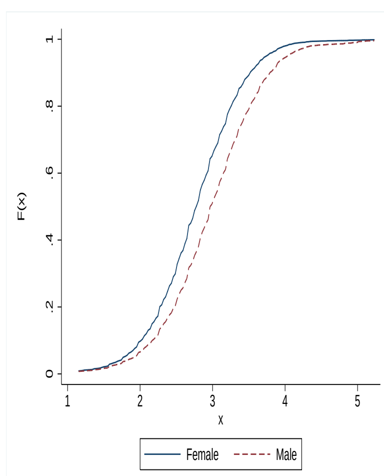
(e) 2006



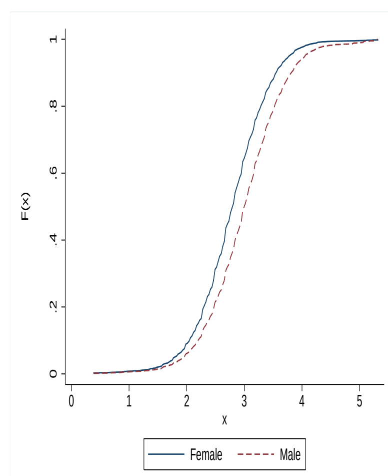
(f) 2007



(g) 2008

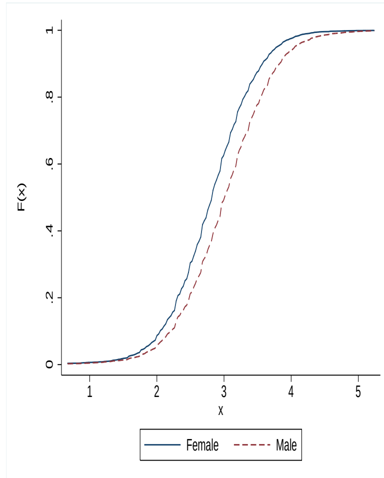


(h) 2009



(i) 2010

Figure 5: CDF Comparisons of Female and Male Wage Distributions (2002 - 2010)



(a) 2011

Figure 6: CDF Comparisons of Female and Male Wage Distributions (2011)

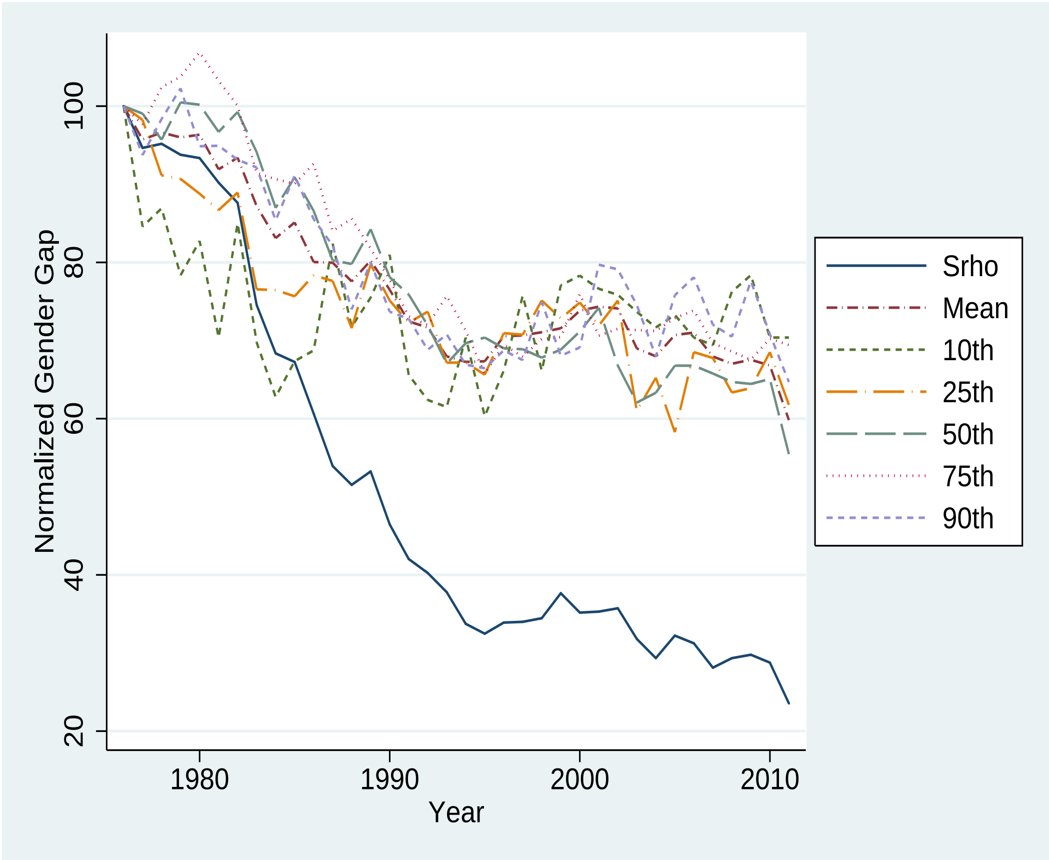
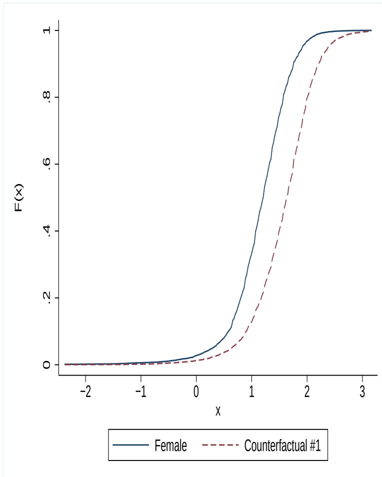
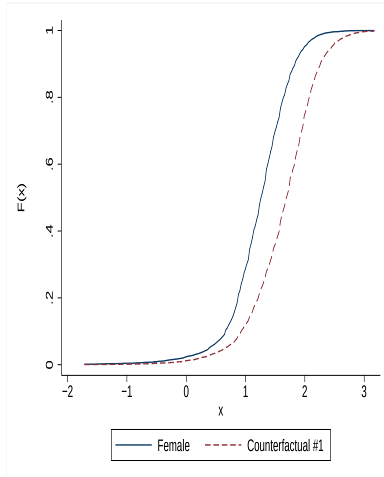


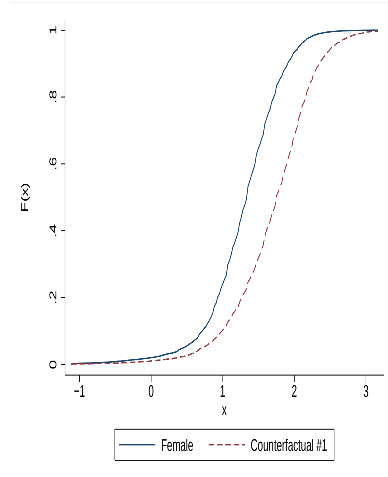
Figure 7: The Trend of the difference between the female wage distribution and the counterfactual distribution #1



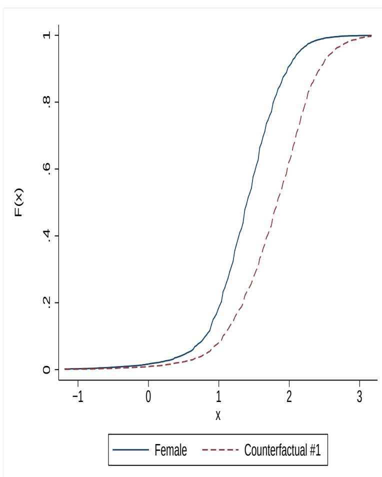
(a) 1976



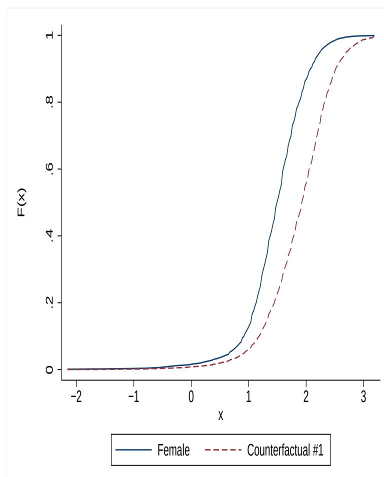
(b) 1977



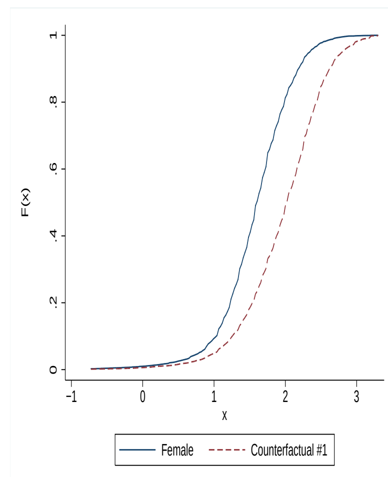
(c) 1978



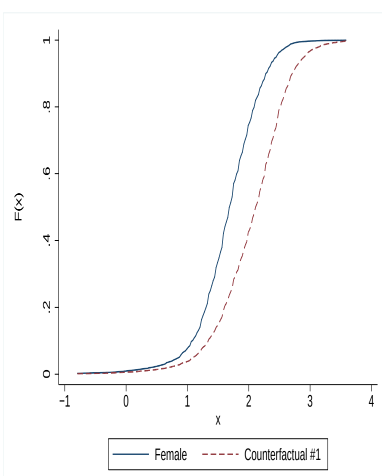
(d) 1979



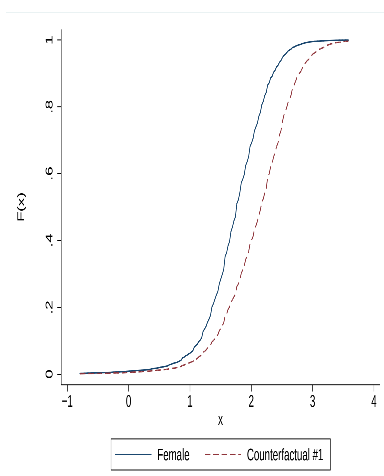
(e) 1980



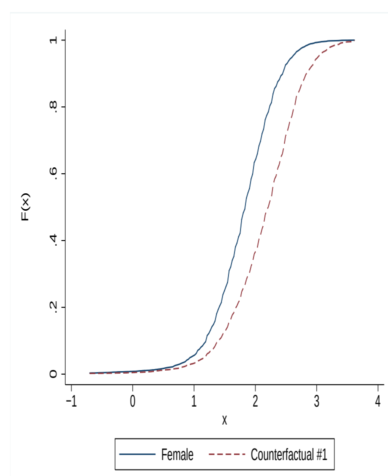
(f) 1981



(g) 1982

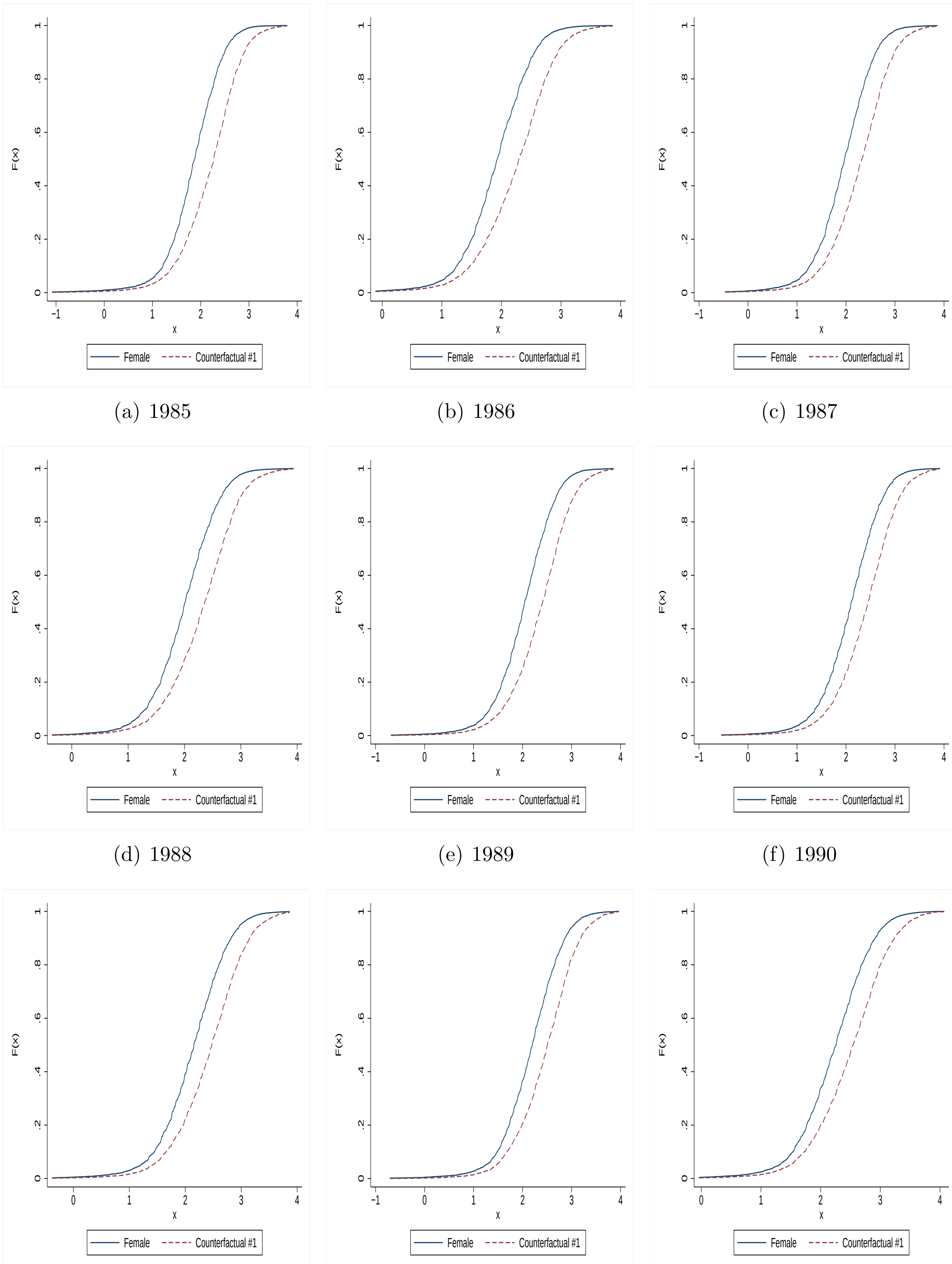


(h) 1983



(i) 1984

Figure 8: CDF Comparisons of Female and Female Counterfactual #1 Wage Distributions (1976-1984)

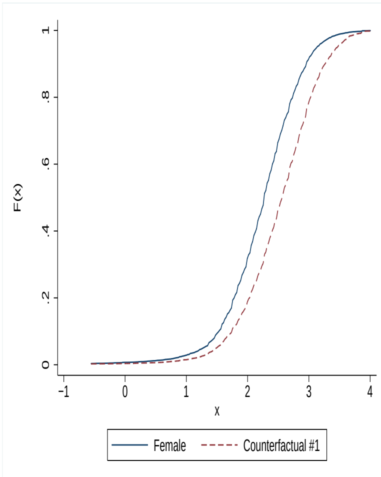


(a) 1985 (b) 1986 (c) 1987

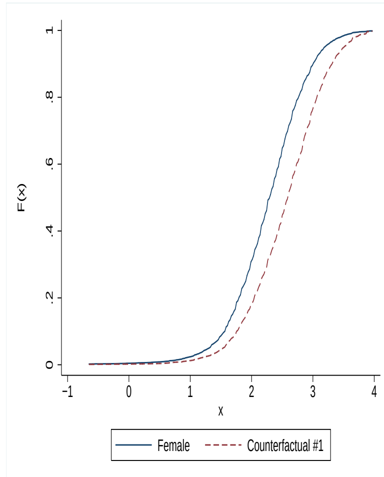
(d) 1988 (e) 1989 (f) 1990

(g) 1991 (h) 1992 (i) 1993

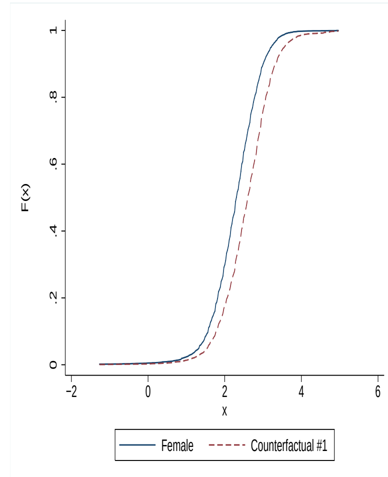
Figure 9: CDF Comparisons of Female and Female Counterfactual #1 Wage Distributions (1985 - 1993)



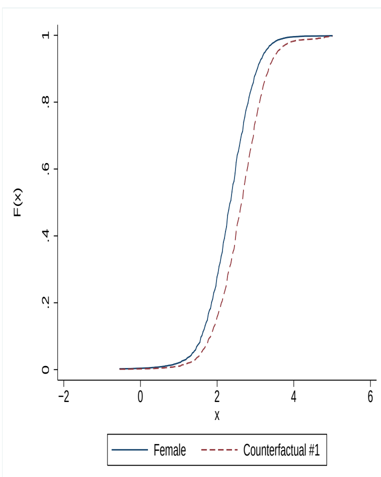
(a) 1994



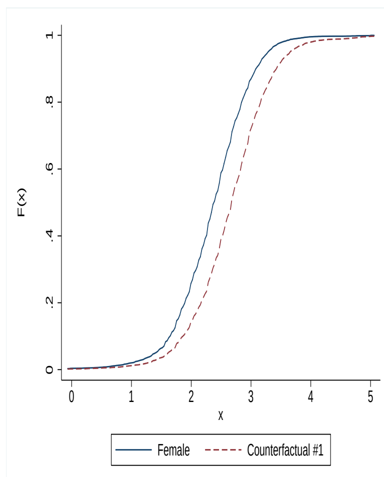
(b) 1995



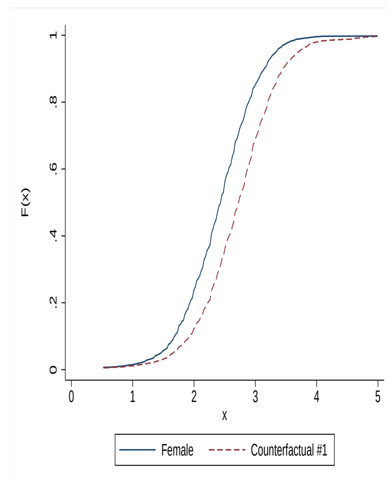
(c) 1996



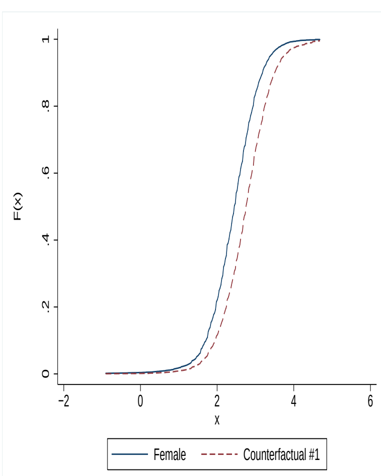
(d) 1997



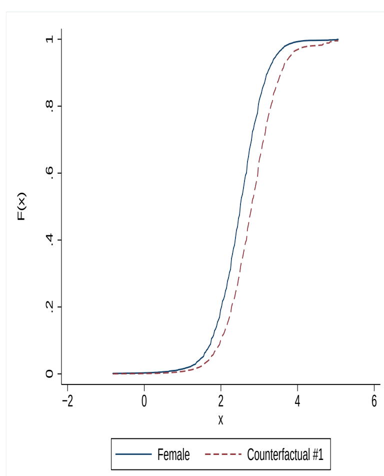
(e) 1998



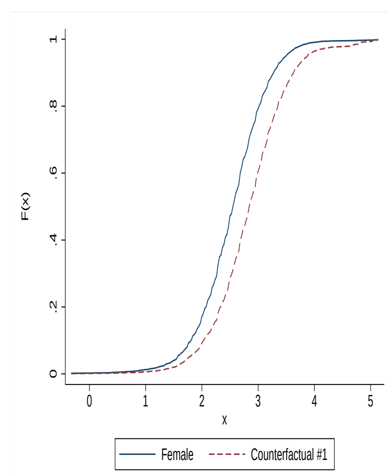
(f) 1999



(g) 2000

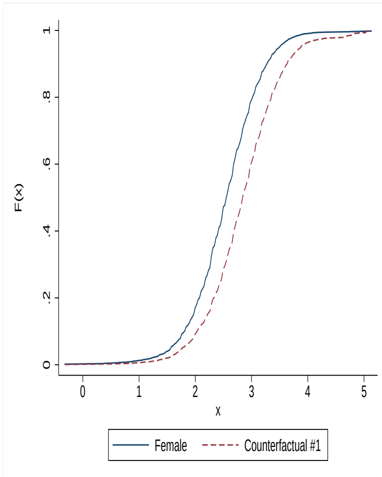


(h) 2001

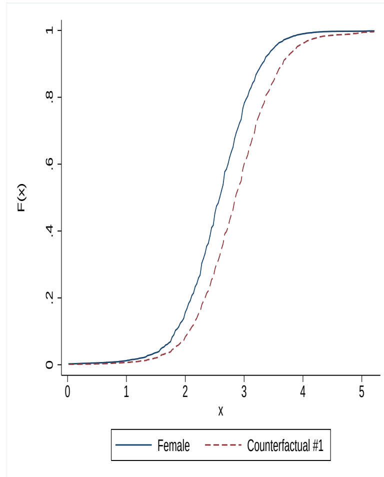


(i) 2002

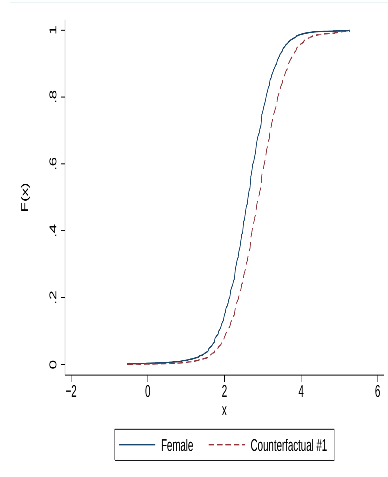
Figure 10: CDF Comparisons of Female and Female Counterfactual #1 Wage Distributions (1994 - 2002)



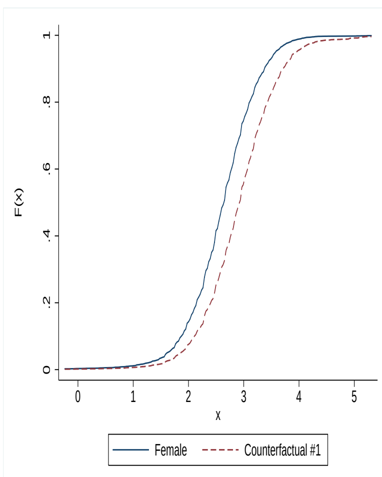
(a) 2002



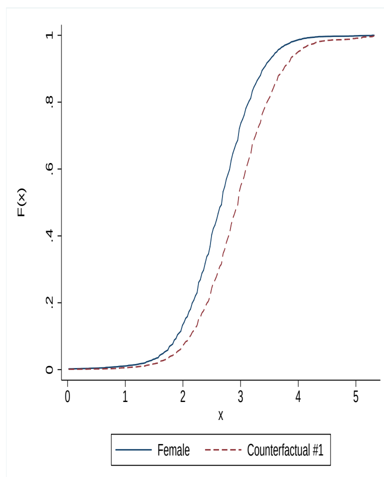
(b) 2003



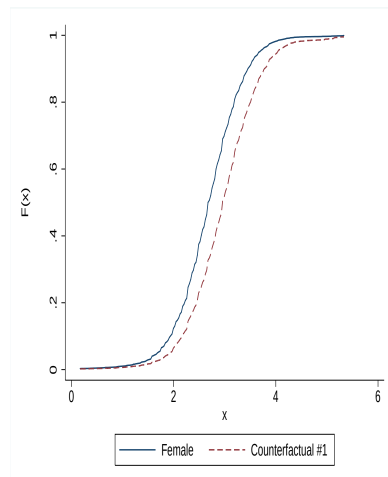
(c) 2004



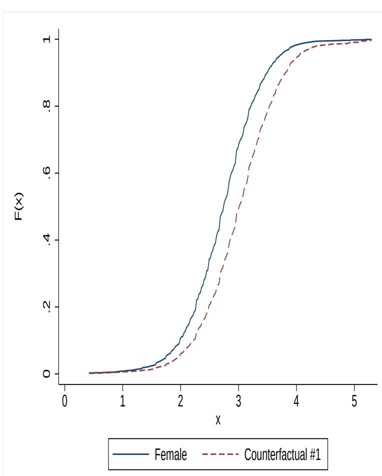
(d) 2005



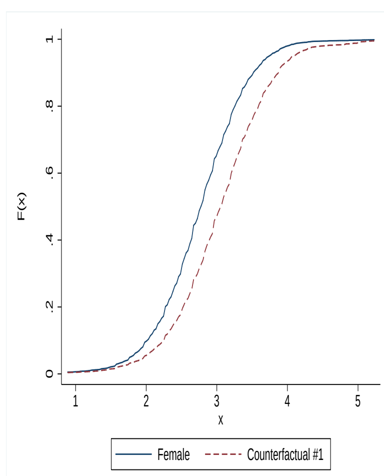
(e) 2006



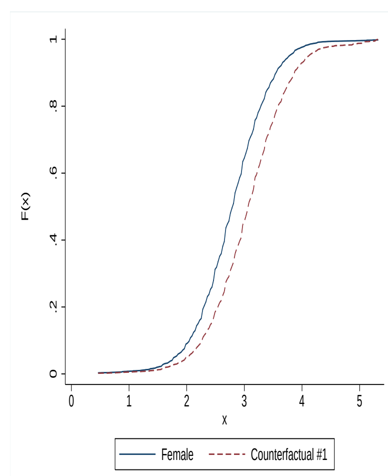
(f) 2007



(g) 2008

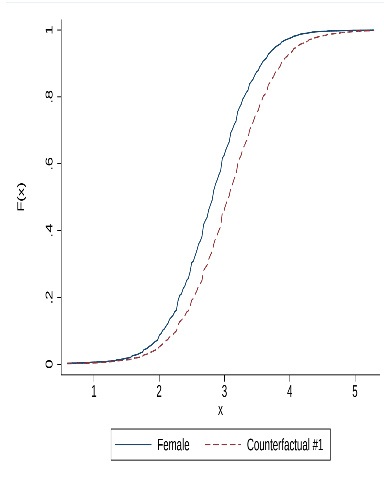


(h) 2009



(i) 2010

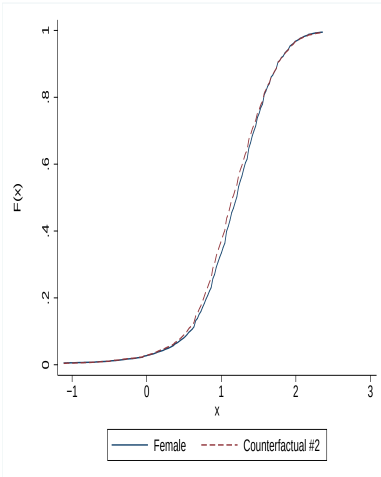
Figure 11: CDF Comparisons of Female and Female Counterfactual #1 Wage Distributions (2002 - 2010)



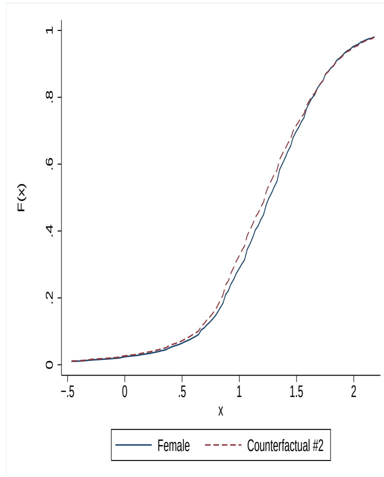
(a) 2011

Figure 12: CDF Comparisons of Female and Female Counterfactual #1 Wage Distributions (2011)

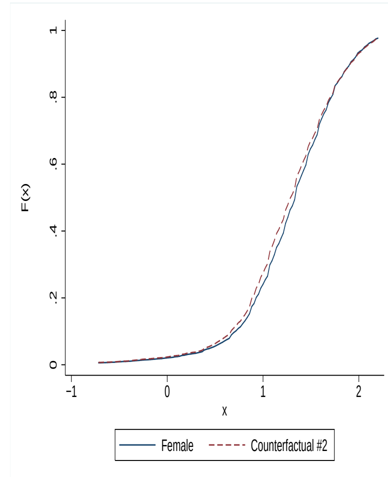
Figure 13: The Trend of the difference between the female wage distribution and the counterfactual distribution #2



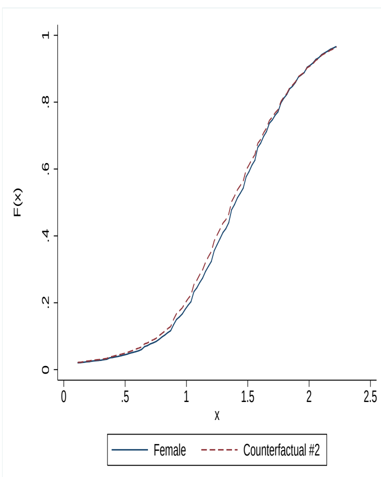
(a) 1976



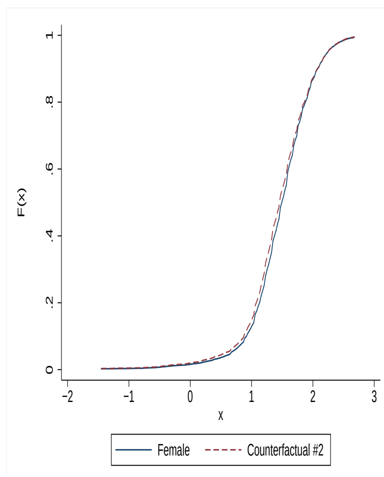
(b) 1977



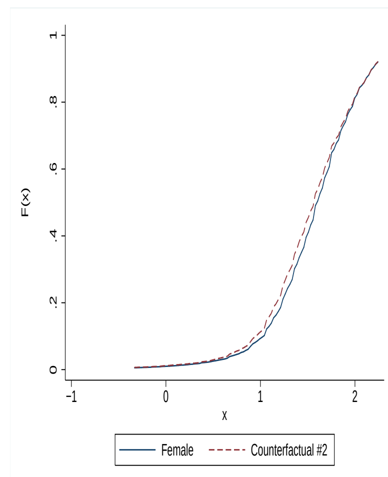
(c) 1978



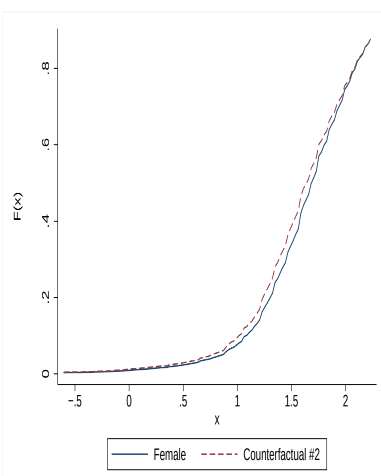
(d) 1979



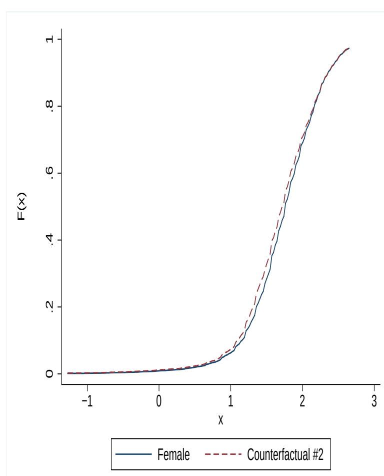
(e) 1980



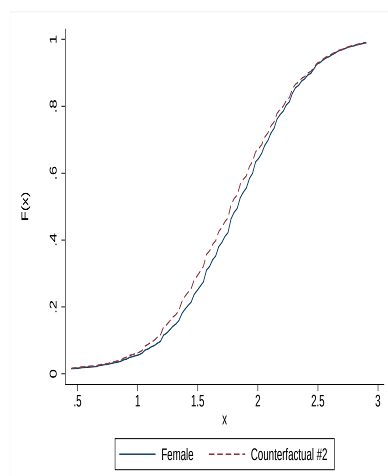
(f) 1981



(g) 1982

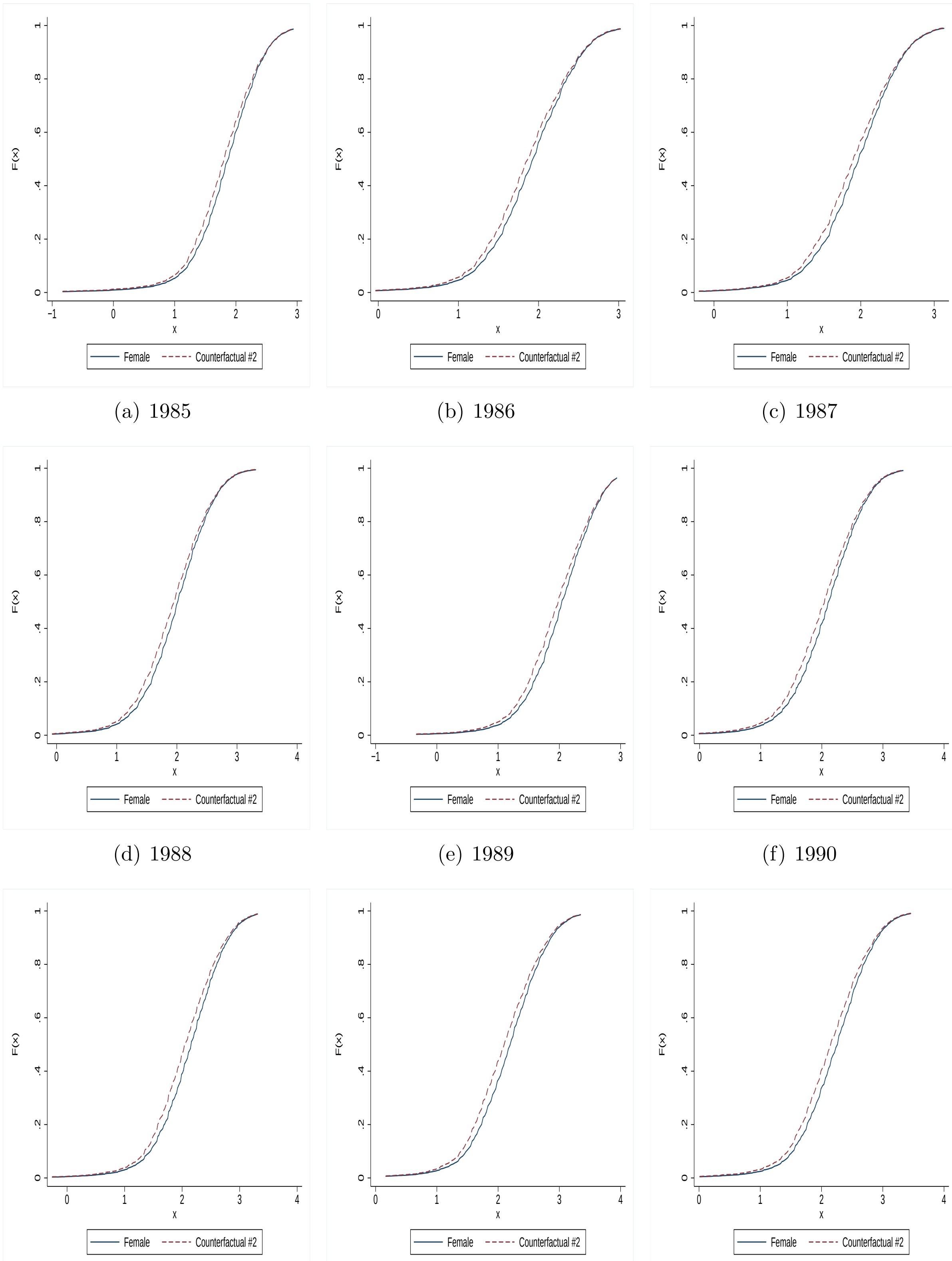


(h) 1983



(i) 1984

Figure 14: CDF Comparisons of Female and Female Counterfactual #2 Wage Distributions (1976-1984)

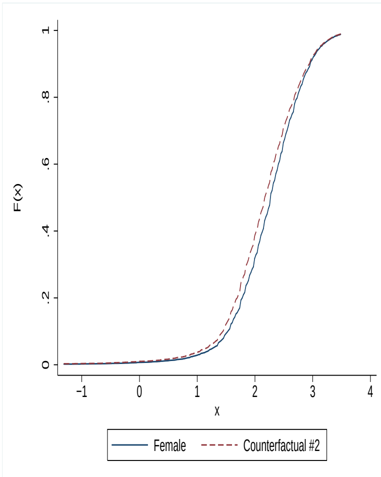


(a) 1985 (b) 1986 (c) 1987

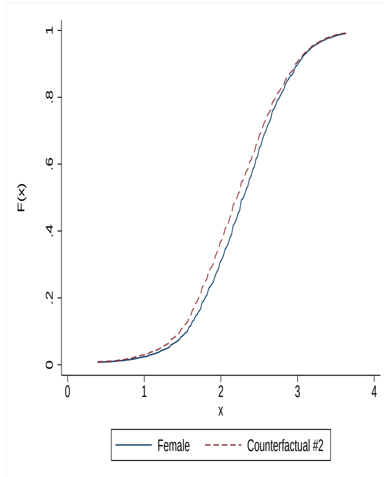
(d) 1988 (e) 1989 (f) 1990

(g) 1991 (h) 1992 (i) 1993

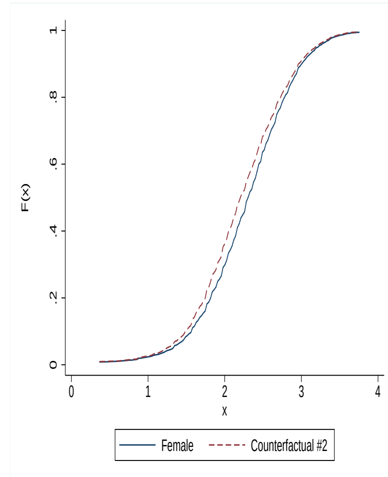
Figure 15: CDF Comparisons of Female and Female Counterfactual #2 Wage Distributions (1985 - 1993)



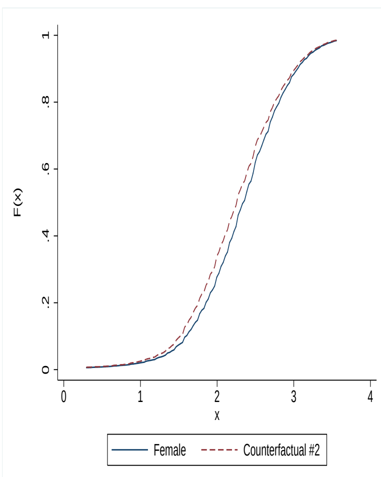
(a) 1994



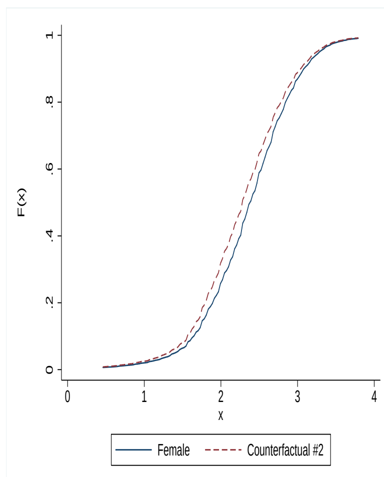
(b) 1995



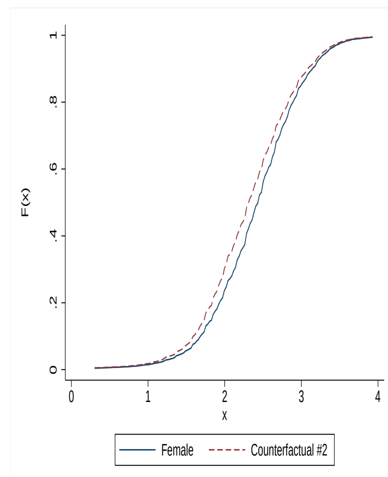
(c) 1996



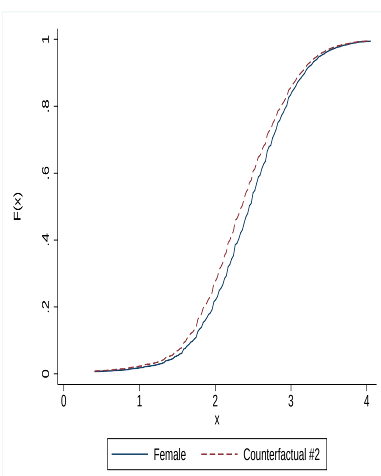
(d) 1997



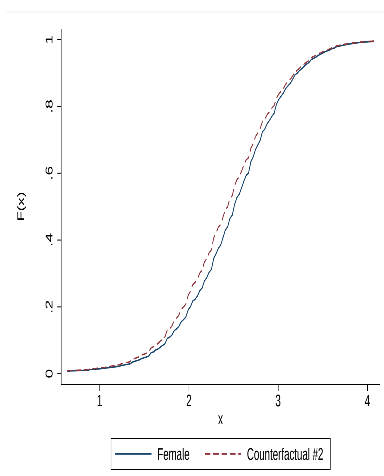
(e) 1998



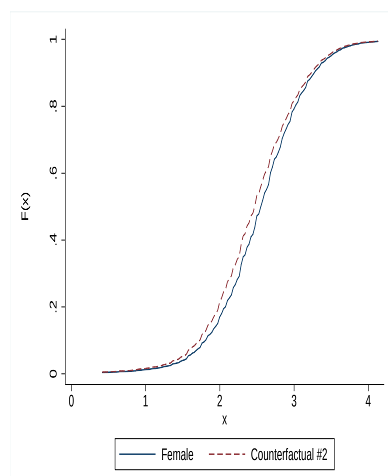
(f) 1999



(g) 2000

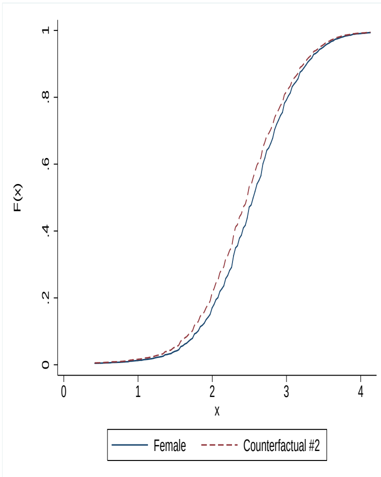


(h) 2001

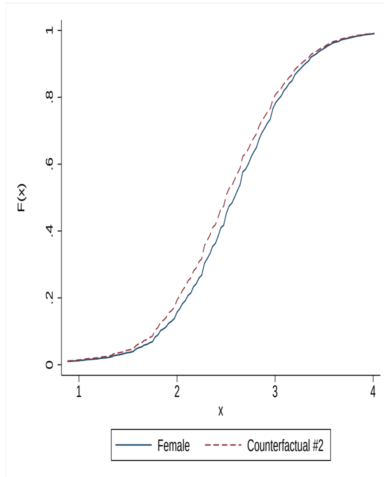


(i) 2002

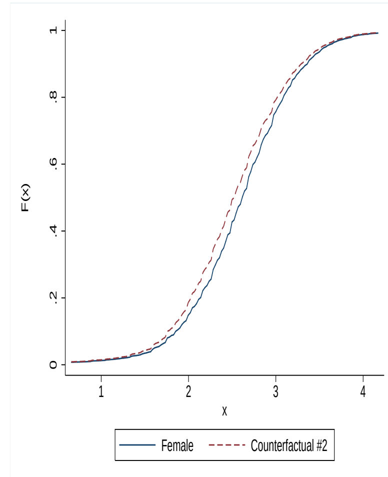
Figure 16: CDF Comparisons of Female and Female Counterfactual #2 Wage Distributions (1994 - 2002)



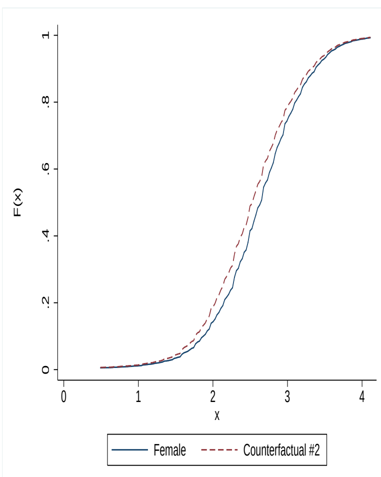
(a) 2002



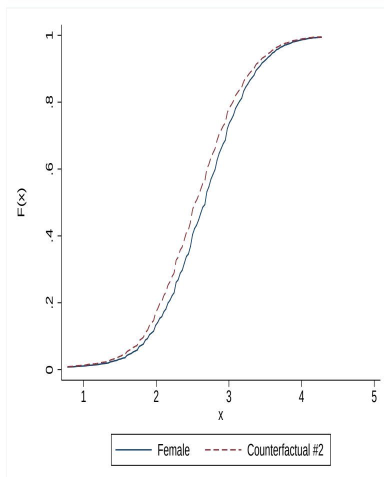
(b) 2003



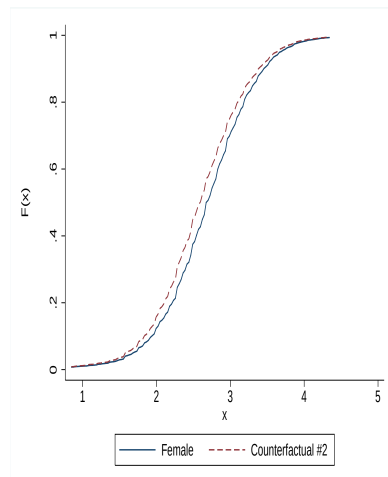
(c) 2004



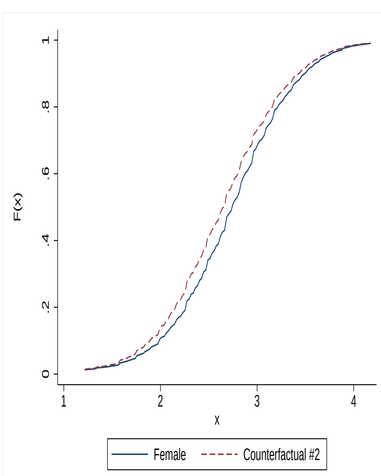
(d) 2005



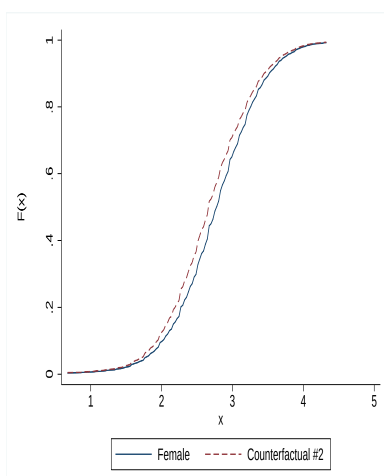
(e) 2006



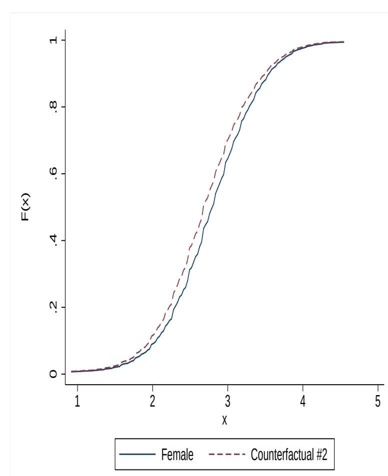
(f) 2007



(g) 2008

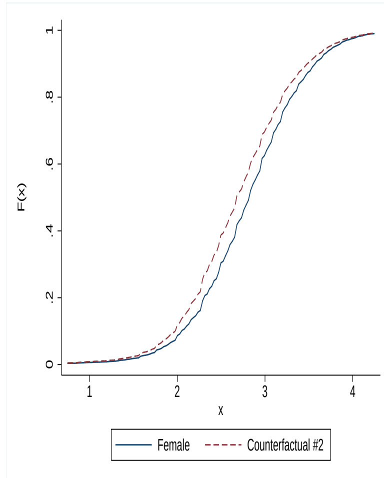


(h) 2009



(i) 2010

Figure 17: CDF Comparisons of Female and Female Counterfactual #2 Wage Distributions (2002 - 2010)



(a) 2011

Figure 18: CDF Comparisons of Female and Female Counterfactual #2 Wage Distributions (2011)

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	D	D
1978	D	I	I	I	D	I	D
1979	D	D	D	D	I	I	I
1980	D	D	D	D	D	D	D
1981	D	D	D	D	D	D	I
1982	D	I	I	I	I	I	D
1983	D	D	D	D	D	D	I
1984	D	D	D	D	D	D	D
1985	D	I	I	D	I	I	I
1986	D	D	I	I	D	D	D
1987	D	D	I	D	D	D	D
1988	D	D	D	D	D	D	D
1989	D	D	I	I	I	D	I
1990	D	D	D	D	D	D	D
1991	D	D	D	I	D	D	D
1992	D	D	D	D	D	D	I
1993	D	D	D	D	D	I	D
1994	D	D	I	D	I	D	D
1995	I	I	I	I	I	I	D
1996	D	I	I	I	D	D	-
1997	D	D	D	D	D	D	D
1998	I	I	D	I	I	D	I
1999	I	D	I	D	D	I	D
2000	D	D	I	D	I	I	D
2001	I	I	I	I	I	I	I
2002	D	D	D	I	D	D	D
2003	D	D	I	D	D	I	D
2004	D	D	D	D	D	D	D
2005	I	D	D	D	D	D	I
2006	D	I	I	I	D	D	D
2007	D	D	D	D	I	D	D
2008	D	D	I	D	D	D	D
2009	I	I	D	I	D	D	I
2010	D	D	I	I	I	I	D
2011	D	D	I	D	D	I	D

[1] Data Source: IPUMS CPS (`http://cps.ipums.org/cps/`). Column (1) reports the overal gender gap ($\times 100$) at corresponding functionals of the distributions of log wages (measures the distance between the female and male wage distributions). Columns (2)- (6) report conventional measures based on difference in parts of the wage distributions between males and females. 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 the measure decreased.

Table A2: Significance Testing of S_ρ

Year	Female v.s. Male			Female v.s. Counterfactual #1			Female v.s. Counterfactual #2		
	90^{th}	95^{th}	99^{th}	90^{th}	95^{th}	99^{th}	90^{th}	95^{th}	99^{th}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1976	0.06	0.07	0.08	0.07	0.08	0.10	0.07	0.08	0.10
1977	0.04	0.05	0.05	0.06	0.06	0.07	0.05	0.06	0.06
1978	0.04	0.05	0.05	0.07	0.07	0.08	0.06	0.07	0.08
1979	0.05	0.05	0.06	0.06	0.06	0.07	0.05	0.05	0.07
1980	0.05	0.05	0.06	0.06	0.06	0.07	0.05	0.05	0.07
1981	0.04	0.04	0.04	0.05	0.05	0.06	0.04	0.04	0.05
1982	0.05	0.05	0.06	0.06	0.07	0.08	0.05	0.06	0.06
1983	0.04	0.05	0.05	0.05	0.06	0.07	0.05	0.05	0.06
1984	0.04	0.05	0.05	0.05	0.06	0.07	0.05	0.05	0.06
1985	0.04	0.05	0.05	0.05	0.05	0.06	0.04	0.05	0.06
1986	0.04	0.05	0.05	0.06	0.07	0.08	0.06	0.06	0.07
1987	0.04	0.05	0.06	0.06	0.06	0.07	0.05	0.05	0.06
1988	0.04	0.04	0.05	0.05	0.05	0.06	0.04	0.04	0.05
1989	0.04	0.05	0.05	0.05	0.06	0.07	0.05	0.05	0.06
1990	0.03	0.04	0.05	0.04	0.05	0.06	0.04	0.04	0.05
1991	0.03	0.04	0.04	0.04	0.04	0.05	0.04	0.04	0.05
1992	0.04	0.04	0.05	0.05	0.05	0.06	0.04	0.04	0.05
1993	0.04	0.04	0.05	0.04	0.05	0.06	0.04	0.04	0.05
1994	0.04	0.05	0.05	0.05	0.06	0.07	0.05	0.06	0.07
1995	0.04	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.05
1996	0.05	0.05	0.06	0.06	0.06	0.07	0.05	0.05	0.06
1997	0.06	0.06	0.07	0.06	0.06	0.07	0.05	0.06	0.07
1998	0.04	0.05	0.06	0.05	0.06	0.07	0.05	0.05	0.07
1999	0.04	0.05	0.06	0.05	0.05	0.06	0.05	0.05	0.06
2000	0.05	0.06	0.07	0.06	0.06	0.07	0.04	0.05	0.06
2001	0.04	0.04	0.04	0.04	0.05	0.05	0.03	0.04	0.04
2002	0.03	0.03	0.03	0.04	0.04	0.04	0.03	0.04	0.04
2003	0.03	0.04	0.04	0.04	0.04	0.05	0.03	0.04	0.04
2004	0.04	0.04	0.04	0.04	0.04	0.05	0.04	0.04	0.05
2005	0.04	0.04	0.04	0.04	0.05	0.05	0.04	0.04	0.05
2006	0.04	0.04	0.04	0.04	0.05	0.05	0.04	0.04	0.05
2007	0.04	0.04	0.04	0.04	0.05	0.05	0.04	0.04	0.05
2008	0.03	0.04	0.04	0.04	0.04	0.05	0.03	0.04	0.04
2009	0.04	0.04	0.05	0.05	0.05	0.06	0.04	0.04	0.05
2010	0.04	0.04	0.05	0.05	0.05	0.06	0.04	0.04	0.05
2011	0.04	0.04	0.04	0.04	0.05	0.05	0.03	0.04	0.04

[1] Data Source: IPUMS CPS (`http://cps.ipums.org/cps/`). Columns (1)-(3) report the 90th, 95th, and 99th percentiles obtained under the null of no difference between male and female wage distributions ($\times$100). Columns (4)-(6) report the 90th, 95th, and 99th percentiles obtained under the null of no difference between female and the counterfactual wage #1 distributions (($\times$100)); Columns (7)-(9) report the 90th, 95th, and 99th percentiles obtained under the null of no difference between female and counterfactual wage #2 distributions (($\times$100)).