CHAPTER 9

WOMEN'S POTENTIAL EARNINGS DISTRIBUTIONS

Esfandiar Maasoumi[a] and Le Wang[b]

[a]*Emory University, Atlanta, GA, USA*
[b]*University of Oklahoma, Norman, OK, USA*

ABSTRACT

Building on recent advances in inverse probability weighted identification and estimation of counterfactual distributions, the authors examine the history of wage earnings for women and their potential wage distributions in the United States. These potentials are two counterfactuals, what if women received men's market "rewards" for their own "skills," and what if they received the women's rewards but for men's characteristics? Using the Current Population Survey data from 1976 to 2013, the authors analyze the entire counterfactual distributions to separate the "structure" and human capital "composition" effect. In contrast to Maasoumi and Wang (2019), the reference outcome in these decompositions is women's observed earnings distribution, and inverse probability methods are employed, rather than the conditional quantile approaches. The authors provide decision theoretic measures of the distance between two distributions, to complement assessments based on mean, median, or particular quantiles. We assess uniform rankings of alternate distributions by tests of stochastic dominance in order to identify evaluations robust to subjective measures. Traditional moment-based measures severely underestimate the declining trend of the structure effect. Nevertheless, dominance rankings suggest that the structure ("discrimination"?) effect is bigger than human capital characteristics.

 Keywords: Gender gap; inequality; entropy; decomposition; discrimination;
 sample selection; treatment effects

1. INTRODUCTION

Both assessment of policy effectiveness and decomposition analysis of between-group differences in outcomes necessarily entail a comparison between two or more potential outcome distributions in different treatment states. For example, when evaluating the effectiveness of the Job Corps programs, America's largest active job market program for at-risk youth, Eren and Ozbeklik (2014) compare the potential earnings distribution when participating in the program to the earnings distribution when not participating. In the analysis of the gender gap, one is often interested in comparing the female earnings distribution to the counterfactual distribution when women and men are "endowed" with the same human capital characteristics, or when women's human capital characteristics are rewarded the same as men in the labor market. The former comparison is typically referenced as "composition effects" (the part of the gender gap due to the differences in market-valued skills and characteristics) and the latter "structural effects" due to the differences in returns to individual characteristics.

While the methodology is different, this chapter follows the distribution-based and evaluative philosophies advocated in Maasoumi and Wang (2019, 2017). Unlike Maasoumi and Wang (2019), we employ more flexible inverse probability weighting methods to identify and estimate distributions and their counterfactuals. And, unlike Maasoumi and Wang (2019), we accommodate "selection" without extensive conditional quantile estimation, by estimating "selection scores" that weight labor market participation. A further distinction of this work is its focus on women's outcome, and its comparison with women's counterfactual outcomes. The gender differences aspect is implicit, and exemplary, not the focus. This is a study of women's potential outcomes in wage distributions.

The most common practice in the literature is to focus on "average" outcomes, with seemingly puzzling deviations at various parts of the wage distribution. In the analysis of the gender gap, the earnings differences between women's average (median) wages and the average (median) of the counterfactual wage distribution are often reported (e.g., Blau & Kahn, 2006; Polachek, 2006). Researchers are increasingly aware of this issue, and, as a result, the differences at other parts of the earnings distributions (e.g., 90*th* percentile) are also reported in recent years.

The conclusions drawn from various measures are, however, often difficult to summarize. In most cases, there are "losers" and "winners" as a result of policy and social change. Overall assessment and ranking of two earnings distributions necessarily entails implicit and subjective weights to different groups. It is important and instructive to be explicit about these subjective assessments.

To make some of our points more concrete, consider the following example for a society with only two groups (Female group A and Female group B). We consider actual earnings and counterfactual earnings. Suppose the difference in group A's earnings and its counterfactual earnings is –$200, and the corresponding number for Female group B is $200. The average difference is $0. No individual experiences this outcome! Policies aimed at dealing with the "average" group/person are not likely to be effective and well targeted. Both quantile effects,

−$200 or $200, are misleading, and any assessment or policy decision would inevitably require subjective weighting of the two groups/individuals.

These dollar differences may occur at different income levels, for example with individual A above $200K (say), and individual B in the $15K range. Averages, medians, and individual quantile differences are seen to implicitly value $1 the same at all income levels and infinite substitutes! Most social and political debates may be seen as a disagreement with this implication of "averages" and similar assessments. Incidentally, this is also an issue with a popular "inequality measure," the difference or ratio between high and low quantiles! Ranking based on any scalar index is subjective and require greater transparency. Given a lack of consensus on any subjective index, robustness testing is a natural complementary assessment.

We first discuss a distributional measure of the gap between two potential outcome distributions based on weights implicit from entropies. One is the normalized Bhattacharay-Matusita-Hellinger entropy, see Granger, Maasoumi, and Racine (2004). The other is a Kullback-Leibler-Theil measure. The latter is symmetrized, but is not a "metric." One important feature, among others, of these measures is their ability to summarize the distance between two *whole* distributions, instead of simple differences between means, medians, or at different *parts* of the distributions. Their welfare theoretic foundations have been explicated in Maasoumi and Wang (2019).

Second, we employ stochastic dominance (SD) tests to assess uniform ranking of wage distributions over entire classes of utility functions. This is similar, but broader than ranking over classes of inequality measures. Inferring a uniform ranking implies that comparisons based on multiple measures is not needed, except when a suitable (cardinal) quantification is desired. An inability to infer a uniform relation is equally informative, indicating that any ranking must be based on a subjective index and its implicit welfare/weighting function.

We first identify the entire distributions of wages and two counterfactual wages among working women. We then perform the comparisons using the proposed assessment approaches. Our comparisons loosely represent two policy scenarios: (1) policies aimed at impacting women's pay structure and (2) policies aimed at impacting observable characteristics and skills.

Using the Current Population Survey (CPS) data 1976–2013 in the United States for our empirical analysis, we reach the following conclusions. First, we find substantial heterogeneity in the implied structure and composition effects across the distribution. Such heterogeneity impacts our perception of the long-run trend of both effects. For example, we find that traditional "centrality" measures severely underestimate the declining trend of the structure effects in the United States. Second, we find first-order stochastic dominance in all cases when comparing the female distribution and the counterfactual distribution when women are endowed with men's wage structure. This result is powerful, suggesting policies aimed at increasing women's pay equity could potentially improve women's welfare *uniformly*. In contrast, in early years, we fail to find any statistically significant dominance relations when comparing the female distribution to the counterfactual distribution when women possess the same distribution of human capital characteristics as men do. In later years, we do find dominance relations, but the results suggest that women's human capital characteristics are

not necessarily inferior to that of men's, and thus policies aimed at changing the human capital characteristics only, may not produce relative improvements for women. Finally, addressing selection impacts primarily the counterfactual distribution when changing the distribution of women's human capital characteristics to that of men's, but the general patterns remain the same.

The rest of the chapter is organized as follows. Section 2 presents the empirical methods employed; Section 3 describes the data; Section 4 discusses the basic results and Section 5 discusses the results addressing selection, and Section 6 concludes.

2. EMPIRICAL METHODS

2.1. Basic Notations

We consider two exclusive outcomes, actual and counterfactual outcomes for women, indexed by $D = d \in \{0,1\}$, with N_d individuals in each group $(N_1 + N_0 = N)$. Let $Y(d)$ be the outcome of interest for any individual in group d (for example, (log) wage *offers*), given by

$$Y(d) = m^d(X,U) \quad \text{for } d \in \{0,1\} \tag{1}$$

where U represents unobservable determinants of wages, X is a vector of observable characteristics distributed from F_X^d defined over the support $\mathcal{X}_d \subset \mathcal{X}$ for each group d. Let $F_{Y(d)|D=d}$ and $F_{Y(d)|D=d,X}$ be the unconditional and conditional distributions of the outcome for group d, respectively.

In practice, we do not observe the full-time wage offers for people who work only part-time or do not work at all; the so-called sample selection issue. Let S be a binary indicator for full-time workers, $S = 1$ if an individual's full-time wages is observed in the sample and zero otherwise. The observed outcome is equal to Y for individuals who are full-time workers ($S = 1$), but missing for those who are not ($S = 0$). Similarly, define the unconditional and conditional distributions of the outcomes for group d in the selected sample, $F_{Y(d)|D=d,S=1}$ and $F_{Y(d)|X,D=d,S=1}$, and the distribution of observable characteristics, $F_{X|S=1}^d \cdot f_{Y(d)|D=d,S=1}$ and $f_{Y(d)|X,D=d,S=1}$ are the corresponding density functions.

Following Maasoumi and Wang (2019), we consider two types of counterfactuals for the target population $T = t$ (which can be all the members in each group or a selected population):

$$F_{C_1|T=t} = \int F_{Y(0)|X,D=1,T=t}(y \mid x,1) dF_{X|T=t}^1(x) \quad \text{(Counterfactual Distribution \#1)}$$
$$F_{C_2|T=t} = \int F_{Y(1)|X,D=0,T=t}(y \mid x,0) dF_{X|T=t}^0(x) \quad \text{(Counterfactual Distribution \#2)} \tag{2}$$

where $F_{C_1|T=t}$ represents the counterfactual distribution when male wage structure, $m^0(X,U)$, is used, holding the distribution of women's human capital characteristics, $F_{X|T=t}^1(x)$ unchanged. $F_{C_2|T=t}$ represents the counterfactual distribution when female wage structure is used, holding the distribution of men's human

capital characteristics unchanged. $F_{g|T=t}$ is the corresponding density function for $F_{g|T=t}, g = C_1, C_2$. The differences in the counterfactual distribution $F_{C_1|T=t}$ and the observed outcome distribution $F_{Y(1)|D=1,T=t}$ provide insight into "structural effects," the differences in wage structure between men and women. The differences in the distributions $F_{C_2|T=t}$ and $F_{Y(1)|D=1,T=t}$ reflect the differences in the distribution of human capital characteristics, the "composition effects." Note the index $T = t$ is dropped when referring to the whole population as the target population.

2.2. Comparing Two Distributions: Entropy-Based Measures and Stochastic Dominance Tests

2.2.1. Entropy-Based Measures

A general definition of the difference between two distributions can be thought of as the difference of respective Evaluation Functions (EFs):

$$EF(F_{Y(1)|D=1,T=t}) - EF(F_{g|T=t}), \quad g = \{C_1, C_2\}$$

Commonly used are the mean difference, the difference in the means of the female earnings distribution and the counterfactual distribution, and the difference at a p_{th} quantile. 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 distances in the female earnings distributions and the counterfactuals.

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

$$KL = \frac{1}{2} \left[\int [\log(\frac{f_1}{f_g}) f_1 + \log(\frac{f_g}{f_1}) f_g] dy \right] \tag{3}$$

2. The Bhattacharya-Matusita-Hellinger measure, given by:

$$S_\rho = \frac{1}{2} \int_{-\infty}^{\infty} \left(f_g^{1/2} - f_1^{1/2} \right)^2 dy \tag{4}$$

where f_1 and f_g the corresponding density functions of $F_{Y(1)|D=1,T=t}$ and $F_{g|T=t}, g = \{C_1, C_2\}$, respectively.

Following Granger et al. (2004) and Maasoumi and Racine (2002), we consider a kernel-based implementation of (3) and (4).[1] The asymptotic distribution of the feasible measures has been derived by Skaug and Tjostheim (1996) and Granger et al. (2004), and are well known to perform very poorly. We employ bootstrap re-sampling procedure based on 299 replications to obtain standard errors for inference.

These entropy measures are founded on certain welfare functions with specific weights to different quantiles. The same is true of the mean and median, associated with the more extreme equal weights and similar welfare functions. Assessments that may be robust to the choice of any welfare function (with a

large class) may be made by tests for stochastic dominance (or prospect domi-
nance, see Linton, Maasoumi, & Whang, 2005). Below, we explicitly introduce
these concepts for (weak) uniform ranking of distributions.

2.2.2. Stochastic Dominance

In the SD approach, the class of social welfare functions underlying the rankings
of the earnings distributions is explicit. Consider two classes of social welfare
functions. U_1 denotes 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., concavity).
Concavity implies an aversion to higher dispersion (or inequality) of wages across
individuals. We are interested in the following scenarios[2]:

Case 1 (First-Order Dominance):
Counterfactual Distribution First-Order Stochastically Dominates Female
Earnings Distribution *if and only if*

1. $\mathbb{E}[u(\ln(w^g))] \geq \mathbb{E}[u(\ln(w^1))]$ for all $u \in U_1$ with strict inequality for some u;
2. Or, $F_{g|T=t}(y) \leq F_{Y(1)|D=1,T=t}(y), g = \{C_1, C_2\}$ for all y with strict inequality for
   some y.

Case 2 (Second-Order Dominance):
Counterfactual Distribution Second-Order Stochastically Dominates Female
Earnings Distribution *if and only if*

1. $\mathbb{E}[u(\ln(w^g))] \geq \mathbb{E}[u(\ln(w^1))]$ for all $u \in U_2$ with strict inequality for some u;
2. Or, $\int_{-\infty}^{y} F_{g|T=t}(t)dt \leq \int_{-\infty}^{y} F_{Y(1)|D=1,T=t}(t)dt, g = \{C_1, C_2\}$ for all y with strict ine-
   quality for some y.

If FSD holds, then the counterfactual earnings distribution is "better" than the
actual female wage distribution for all policymakers with increasing utility func-
tions in the class U_1 (with strict inequality holding for some welfare function(s) in
the class), since the expected social welfare from the counterfactual state is larger
or equal to that from the actual female wage distribution. Note that FSD implies
that the average counterfactual wages are greater than the average actual wages.
"However, a ranking of the average wages does not imply that one FSD the other;
rather, the entire distribution matters" (Mas-Colell, Whinston, & Green, 1995,
p. 196). Similarly, if SSD holds, the counterfactual earnings distribution is "bet-
ter" than the actual wage distribution for those with 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 our proposed approach is that our conclusions do not depend on any specific
functions or weights assigned to the distributions. 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.

In this chapter, 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 given by

$$d = \sqrt{\frac{N_0 N_1}{N_0 + N_1}} \min \sup [F_{g|T=t}(y) - F_{Y(1)|D=1,T=t}(y)] \qquad (5)$$

$$s = \sqrt{\frac{N_0 N_1}{N_0 + N_1}} \min \sup \int_{-\infty}^{y} [F_{g|T=t}(t) - F_{Y(1)|D=1,T=t}(y)]dt, \quad g = \{C_1, C_2\} \qquad (6)$$

Practical implementation of these test statistics is based on the sample counterparts of d and s by replacing CDFs with empirical ones.[3] The underlying distribution of the test statistics are generally unknown and depend on the data. Following the literature (e.g., Maasoumi & Heshmati, 2000; Millimet & Wang, 2006), we use simple bootstrap technique based on 299 replications to obtain the empirical distribution of the test statistics. If the probability of d lies in the nonpositive interval (i.e., $Pr[d \leq 0]$ is large, say 0.90 or higher, and $\hat{d} \leq 0$, we can infer FSD to a desirable degree of statistical confidence. We now turn to identification and estimation of the CDFs.

2.3. Identification and Estimation of Counterfactual Distributions

2.3.1. Without Selection

To identify the counterfactual distributions in (2), a key assumption is the availability of a vector of human capital characteristics for each individual, X, such that the distribution of the unobservables such as ability is independent of the individual state, conditional of X. The assumption permits a causal interpretation of the difference between the earnings and counterfactual distributions. Formally,

Assumption 1 (Ignorability or Conditional Independence Assumption). Let (D, X, U) have a joint distribution. For all x in the support of $\mathcal{X}: U \perp D \mid X = x$.

Moreover, there should also be an overlap in observable characteristics between the two states, or $\mathcal{X}_1 \subset \mathcal{X}_0$ for the integral in Equation (2) to be well-defined. Formally,

Assumption 2 (Overlapping Support). For all x in the support of $\mathcal{X}: 0 < \Pr[D=1 \mid X = x] < 1$ and $\Pr[D=1] > 0$.

As shown in Fortin, Lemieux, and Firpo (2011) and Firpo, Fortin, and Lemieux (2007), these assumptions are sufficient to identify the wage distributions for both states, as well as the counterfactual distributions of interest.

Proposition 1 (Theorem 1 in Firpo et al. (2007): Inverse Probability Weighting).

Under Assumptions (1) and (2)

1. Observed Outcome Distributions: $F_{Y(d)}(y) = \mathbb{E}[\omega_d(D)\, \mathbf{1}[Y \leq y]]$, $d = (0,1)$
2. Counterfactual Outcome Distributions: $F_{C_1}(y) = \mathbb{E}[\omega_C(D, X)\, \mathbf{1}[Y \leq y]]$

The corresponding weighting functions are given by

$$\omega_1 = \frac{D}{p}, \omega_0 = \frac{1-D}{1-p}, \omega_C = \frac{1-D}{p}\frac{\pi(X)}{1-\pi(X)}$$

where $p = \Pr[D=1]$ and $\pi(X) = \Pr[D=1 \,|\, X]$. Note that $F_{C_2}(y)$ can be similarly obtained.

2.3.2. Accounting for Selection

In the presence of sample selection, however, *Assumption 1* may fail to hold in the selected sample, and further assumptions are required for identification of the counterfactual distributions. Huber (2014) and Maasoumi and Wang (2017) propose the following assumption regarding the selection mechanism and availability of an exclusion restriction:

Assumption 3 (Selection Mechanism). The selection mechanism is given by

$$S = \mathbf{1}[V \leq \Pi(X,Z)] \tag{7}$$

where $\Pi(\ )$ is an unknown function, V is an unobservable error term that could be correlated with U, and its distribution, $F_V(v)$ is strictly monotonic.[4] Z is an exclusion restriction that satisfies the following conditions stated in Assumption 4.

Assumption 4 (Exclusion Restriction).

1. Existence of Correlation) $\mathbb{E}[Z\,S\,|\,X] \neq 0$
2. (Conditional Independence) $(U,V) \perp (D,Z) \,|\, X$[5]

Furthermore, similar to *Assumption 2*, we also require that "state/gender" cannot be perfectly predicted by these variables in the selected sample. Formally,

Assumption 5 (Overlapping Support for the Selected Sample). For all $x, p(w)$ in the support of $\mathcal{X} \times P$, $0 < \Pr[D=1 \,|\, X=x, p(W)=p(w), S=1] < 1$, and $\Pr[S=1 \,|\, D=d] > 0$, where $d = \{0,1\}$, where $W \equiv (X,Z)$ and $p(W) \equiv \Pr[S=1 \,|\, X,Z] = F_V(\Pi(X,Z))$; the selection propensity score.

As shown in Huber (2014) and Maasoumi and Wang (2017), we can identify the counterfactual distributions for the selected group under these assumptions.

Proposition 2 (Inverse Probability Weighting). Under *Assumptions 3–5*,

1. **Observed Outcome Distribution**:
   $F_{Y(d)|D=d,S}(y \,|\, d,s) = \mathbb{E}[\omega_d(D)\,\mathbf{1}[Y \leq y] \,|\, S=1], d = (0,1)$
2. **Counterfactual Outcome Distribution**:
   $F_{C_1|S}(y \,|\, s) = \mathbb{E}[\omega_C(D,W)\,\mathbf{1}[Y \leq y] \,|\, S=1]$

The corresponding weighting functions are given by

$$\omega_1 = \frac{D}{p}, \omega_0 = \frac{1-D}{1-p}, \omega_{C_1} = \frac{1-D}{p}\frac{\pi(X,p(W))}{1-\pi(X,p(W))}$$

where $p = \Pr[D = 1 \mid S = 1]$ and $\pi(X, p(W)) = \Pr[D = 1 \mid X, p(W), S = 1]$.

We follow the four-step procedure described in Maasoumi and Wang (2017) to construct the weights and the counterfactual distributions. First, we estimate the logit model of S on X and Z to obtain the estimates of propensity scores, $\widehat{p(W)}$.[6] Second, we obtain estimates of $\pi(X, p(W))$ with the predicted values using a logit model of D on $X, \widehat{W}$. Finally, we obtain the distributional features with the reweighted samples based on the weighting function estimated using their normalized sample analogs,[7] All standard errors are obtained via bootstrapping.

3. DATA

To perform our analysis, we use data from the 1976–2013 March Current Population Survey (CPS) (available at http://cps.ipums.org, Flood, King, Ruggles, & Robert Warren, 2017). The March CPS is a large nationally representative household data that contain detailed information on labor market outcomes and demographic characteristics needed for study of the gender gap (e.g., Mulligan & Rubinstein, 2008; Waldfogel & Mayer, 2000) and our counterfactual analysis. We closely follow Maasoumi and Wang (2019) to construct our variables and samples, and hence provide limited details here. 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 full-time workers with stronger attachment to the labor market – those who worked for more than 20 weeks (inclusive) and more than 35 hours per week in the previous year.

Following the literature (e.g., Blau & 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. We exclude extremely low wages that are less than one unit of the log wages. We also drop imputed wages since the literature has shown inclusion of these values to be problematic and recommended exclusion of these observations (e.g., Bollinger, & Hirsch, 2013). The differences in a specific part of the distribution (such as median) can be interpreted as percentage differences. Note, however, that our distributional measure of the differences and SD tests are invariant to increasing monotonic transformation, while conventional measures are.

In our counterfactual analysis, we include age and its polynomial terms up to fourth order, years of schooling and its square, dummy variables for current marital status, 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). In estimating propensity scores, we also include interaction terms between continuous variables and dummy variables.

Following much of the literature, we use as exclusion restriction in the selection equation whether there is a child under age five in the household. For example,

Mulligan and Rubinstein (2008) use the number of children younger than six, interacted with marital status as variables determining employment, but excluded from the wage equation. While the empirical validity of this variable continues to be debated, it is theoretically clear as to why it is a good candidate,[8] and much of the literature has provided empirical evidence supporting the validity of the exclusion restriction in this context (e.g., Huber & Mellace, 2014) and using the same data as ours (Maasoumi & Wang, 2019). This is indeed the tradition that we follow here. In estimating selection propensity scores, we also include interaction terms between the exclusion restriction and all human capital characteristics.

4. BASELINE RESULTS

4.1. Female Wage Versus Counterfactual Distribution #1

4.1.1. Entropy and Conventional Measures of the Differences

Table 1 reports various measures of the differences between the female wage distribution and the counterfactual wage distribution (#1) (i.e., the distribution of women's wages when their human capital characteristics are paid under men's wage structure). Columns (1) and (2) report our distributional measures of the differences between the two distributions. Note that both measures are normalized, taking on values in [0,1], and to facilitate the presentation, the results reported are the original values $\times 100$ throughout the chapter. Columns (3)–(8) display the difference measured at select percentiles of the distributions (mean, 10th, 25th, 50th, 75th, and 90th) that are commonly used in the literature. The standard errors based on 299 replications are reported in the Online Appendix.

We first notice that all measures imply that there exist substantial differences between female wages and counterfactual wages. In particular, both S_ρ and Theil measures are statistically different from zero. Furthermore, examination of the differences at the select percentiles of the female wage distribution and the counterfactual distribution are consistently positive. Recall that the difference captures only the difference in wage structures between men and women, while holding women's human capital characteristics constant. Therefore, this result indicates that had women been rewarded the same as men in the labor market, they could have higher wages. 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.

However, the implied size of the structure effect (and the potential policy impact) varies with the conventional measure used, with the differences generally smaller at the lower tail of the distribution. For example, in 1976, the measure at the 10th percentile indicates the structure effect is roughly 26 percent, while the measure at the percentiles above the median implies that it is more than 40 percent. If this difference is interpreted as "potential discrimination" as the literature typically does, high-skilled women may face even more discrimination than low-skilled women, suggesting the glass-ceiling effect.

Heterogeneity in the implied structure effect may not necessarily be a problem in the cross-sectional setting. It may, however drastically mask the long-run trend

Table 1. Female Wage Distribution Versus Counterfactual Distribution #1 (Without Selection Correction): Structural Effects.[a]

Year	srho	theil	mean	qte10	qte25	qte50	qte75	qte90
1976	8.689	18.169	0.384	0.275	0.376	0.410	0.405	0.408
1977	8.349	17.389	0.375	0.268	0.372	0.402	0.407	0.405
1978	8.509	17.775	0.386	0.278	0.353	0.405	0.427	0.405
1979	8.196	17.055	0.375	0.222	0.350	0.422	0.437	0.427
1980	8.294	17.224	0.376	0.255	0.357	0.410	0.440	0.399
1981	8.101	16.947	0.371	0.258	0.331	0.411	0.429	0.414
1982	7.870	16.435	0.372	0.248	0.353	0.404	0.437	0.395
1983	6.701	13.869	0.350	0.246	0.309	0.397	0.386	0.392
1984	6.022	12.506	0.336	0.239	0.291	0.379	0.389	0.377
1985	5.843	12.114	0.337	0.250	0.288	0.369	0.377	0.401
1986	5.231	10.838	0.319	0.222	0.291	0.357	0.378	0.368
1987	4.815	9.853	0.317	0.238	0.300	0.353	0.342	0.343
1988	4.574	9.372	0.311	0.240	0.297	0.324	0.360	0.336
1989	4.662	9.911	0.316	0.248	0.298	0.346	0.335	0.350
1990	4.101	8.401	0.299	0.220	0.288	0.316	0.322	0.309
1991	3.771	7.685	0.286	0.222	0.282	0.327	0.307	0.316
1992	3.599	7.328	0.281	0.203	0.282	0.305	0.315	0.308
1993	3.302	6.721	0.265	0.182	0.252	0.288	0.303	0.303
1994	3.028	6.142	0.263	0.214	0.248	0.291	0.291	0.297
1995	2.928	5.931	0.265	0.190	0.254	0.288	0.288	0.291
1996	2.982	6.227	0.273	0.223	0.254	0.288	0.280	0.280
1997	3.134	6.389	0.282	0.222	0.288	0.301	0.289	0.309
1998	3.060	6.213	0.280	0.207	0.273	0.266	0.288	0.310
1999	3.356	6.883	0.289	0.233	0.292	0.288	0.288	0.293
2000	2.967	6.021	0.280	0.223	0.264	0.275	0.300	0.288
2001	3.086	6.287	0.288	0.232	0.257	0.293	0.292	0.339
2002	3.121	6.357	0.290	0.232	0.255	0.280	0.297	0.334
2003	2.691	5.486	0.266	0.204	0.220	0.251	0.288	0.326
2004	2.524	5.114	0.260	0.214	0.223	0.259	0.279	0.288
2005	2.696	5.483	0.273	0.237	0.247	0.259	0.301	0.328
2006	2.669	5.420	0.275	0.207	0.257	0.288	0.303	0.327
2007	2.452	4.959	0.269	0.236	0.257	0.255	0.272	0.306
2008	2.557	5.186	0.267	0.230	0.266	0.268	0.277	0.294
2009	2.496	5.055	0.266	0.223	0.256	0.262	0.280	0.330
2010	2.426	4.912	0.265	0.204	0.228	0.272	0.288	0.318
2011	2.347	4.747	0.260	0.177	0.223	0.251	0.302	0.289
2012	2.326	4.699	0.262	0.182	0.207	0.283	0.297	0.288
2013	2.143	4.329	0.255	0.192	0.203	0.280	0.288	0.262

[a]Data Source: IPUMS CPS (http://cps.ipums.org/cps/). Columns (1) and (2) report the entropy gap measures ($\times 100$) at corresponding functionals of the distributions of log wages (measures the distance between the female and counterfactual wage). Columns (3)–(8) report conventional measures based on difference in parts of the wage distributions between the female wage distribution and the counterfactual distribution.

in the structure effect for the entire society. This is especially true should our goal is to characterize the potential overall impact of a policy aimed at improving market structure for women. As we can see from the table, the implied importance of structural difference over time also varies across measures. While there appears to be a declining trend in the difference at the upper tail, the difference in the lower

tail remains rather stable. Such stark contrast is even more pronounced during the pre-welfare reform period (before 1994). The average rate of change is about 2 percent at the percentiles above 25th percentile, compared to less than 0.1 percent at the 10th percentile. Our entropy measures become particularly useful in this case when the commonly used measures disagree with each other. Our measures summarize the information that a measure at a specific part of the distribution misses. In fact, the rate of decline implied by our entropy measures is much larger than that by the conventional measures. All traditional measures appear to severely *underestimate* the decline in the importance of structure effect over time. In particular, both of our entropy measures imply the structure effect decreases at average annual rate of about 3.5 percent. Intuitively, this makes sense. If the difference at every part of the earnings distributions decreases, the decrease in the distance between the female distribution and the counterfactual distribution should be even larger. This property is also evident in Maasoumi and Wang (2019) when contrasting the entropy measures to the conventional measures when measuring the gender gap itself.

4.1.2. Stochastic Dominance Test Results

As discussed above, these measures of the gender gap do not lend themselves to ranking of the distributions. Therefore, we now to turn to SD tests. SD results are reported in Table 2 and the corresponding comparisons of CDFs for select years plotted in Fig. 1 (the full set of results are available in the Online Appendix). 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 counterfactual distribution lies predominantly to the right of the earnings distribution among women. This casual observation is consistent with the fact that the differences in selected percentiles of the female wage and counterfactual distributions are uniformly positive. According to the actual SD test statistics, we find first-order dominance relations in all cases, and such observed rankings are statistically significant. This result again indicates that women could have been *uniformly* better should their human capital characteristics are rewarded the same in the labor market. As noted in Maasoumi and Wang (2019), such results point to such policies as equity pay as potentially policy candidates to closing the gender gap (e.g. Gunderson & Riddell, 1992; Hartmann & Aaronson, 1994). Such results are even stronger than what is implied by various measures of the gap above.

4.2. Female Wage Versus Counterfactual Distribution #2

Table 3 reports various measures of the differences between the female wage distribution and the counterfactual wage distribution (#2) (i.e., the distribution of women's wages when they possess men's human capital characteristics but

Table 2. Stochastic Dominance Results Without Selection Correction.

Year	Observed Ranking	d	$Pr[d \leq 0]$	s	$Pr[s \leq 0]$	Observed Ranking	d	$Pr[d \leq 0]$	s	$Pr[s \leq 0]$
	Panel A: vs Counter factual #1					**Panel B: vs Counter factual #2**				
1976	FSD	−0.646	1.000	−0.689	1.000	FSD	−0.647	0.441	−0.704	1.000
1977	FSD	−0.699	1.000	−0.744	1.000	FSD	−0.750	0.458	−0.764	1.000
1978	FSD	−0.763	1.000	−0.763	1.000	SSD	0.838	0.324	−0.736	1.000
1979	FSD	−0.721	1.000	−0.804	1.000	FSD	−0.712	0.441	−0.712	1.000
1980	FSD	−0.780	1.000	−0.790	1.000	FSD	−0.799	0.542	−0.799	1.000
1981	FSD	−0.824	1.000	−0.833	1.000	FSD	−0.809	0.699	−1.061	1.000
1982	FSD	−0.729	1.000	−0.767	1.000	FSD	−0.757	0.609	−0.813	1.000
1983	FSD	−0.737	1.000	−0.906	1.000	SSD	1.188	0.090	−0.729	1.000
1984	FSD	−0.721	1.000	−0.724	1.000	FSD	−0.584	0.783	−0.917	1.000
1985	FSD	−0.735	1.000	−0.746	1.000	FSD	−0.820	0.592	−0.909	1.000
1986	FSD	−0.814	1.000	−0.909	1.000	FSD	−0.756	0.997	−0.841	1.000
1987	FSD	−0.736	1.000	−0.736	1.000	FSD	−0.748	0.953	−0.755	1.000
1988	FSD	−0.791	1.000	−1.140	1.000	FSD	−0.825	0.997	−0.877	1.000
1989	FSD	−0.818	1.000	−0.877	1.000	FSD	−0.771	0.916	−0.811	1.000
1990	FSD	−0.900	1.000	−0.901	1.000	FSD	−0.853	1.000	−0.853	1.000
1991	FSD	−0.911	1.000	−0.911	1.000	FSD	−0.850	1.000	−0.945	1.000
1992	FSD	−0.815	1.000	−0.815	1.000	FSD	−0.793	1.000	−1.173	1.000
1993	FSD	−0.813	1.000	−0.907	1.000	FSD	−0.806	1.000	−0.834	1.000
1994	FSD	−0.799	1.000	−0.897	1.000	FSD	−0.897	1.000	−0.932	1.000
1995	FSD	−0.835	1.000	−0.835	1.000	FSD	−0.823	1.000	−0.823	1.000
1996	FSD	−0.795	1.000	−0.812	1.000	FSD	−0.744	1.000	−0.791	1.000
1997	FSD	−0.749	1.000	−0.749	1.000	FSD	−0.768	1.000	−0.774	1.000
1998	FSD	−0.726	1.000	−0.750	1.000	FSD	−0.748	1.000	−0.755	1.000
1999	FSD	−0.844	1.000	−1.012	1.000	FSD	−0.780	1.000	−0.780	1.000
2000	FSD	−0.777	1.000	−0.777	1.000	FSD	−0.772	1.000	−0.827	1.000
2001	FSD	−1.066	1.000	−1.066	1.000	FSD	−1.003	1.000	−1.170	1.000
2002	FSD	−0.965	1.000	−0.989	1.000	FSD	−1.039	1.000	−1.108	1.000
2003	FSD	−0.968	1.000	−1.026	1.000	FSD	−0.957	1.000	−0.957	1.000
2004	FSD	−1.007	1.000	−1.007	1.000	FSD	−0.958	1.000	−0.958	1.000
2005	FSD	−0.987	1.000	−0.987	1.000	FSD	−0.943	1.000	−0.943	1.000
2006	FSD	−0.977	1.000	−0.977	1.000	FSD	−0.940	1.000	−0.986	1.000
2007	FSD	−0.958	1.000	−0.958	1.000	FSD	−0.948	1.000	−0.955	1.000
2008	FSD	−0.963	1.000	−0.963	1.000	FSD	−0.942	1.000	−1.400	1.000
2009	FSD	−1.015	1.000	−1.054	1.000	FSD	−0.966	1.000	−1.050	1.000
2010	FSD	−0.956	1.000	−0.966	1.000	FSD	−0.934	1.000	−0.972	1.000
2011	FSD	−0.906	1.000	−1.183	1.000	FSD	−0.936	1.000	−0.936	1.000
2012	FSD	−0.958	1.000	−1.013	1.000	FSD	−0.940	1.000	−0.998	1.000
2013	FSD	−0.905	1.000	−0.952	1.000	FSD	−0.912	1.000	−0.969	1.000'

holding women's wage structure unchanged). 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 less than a half of the structural effect, and in some case at the upper tail (before 1990 at the 90th percentile), the magnitude is

(a) 1976

(b) 1980

(c) 1990

(d) 2000

(e) 2010

(f) 2013

vs. Counterfactual # 1

(g) 1976

(h) 1980

(i) 1990

(j) 2000
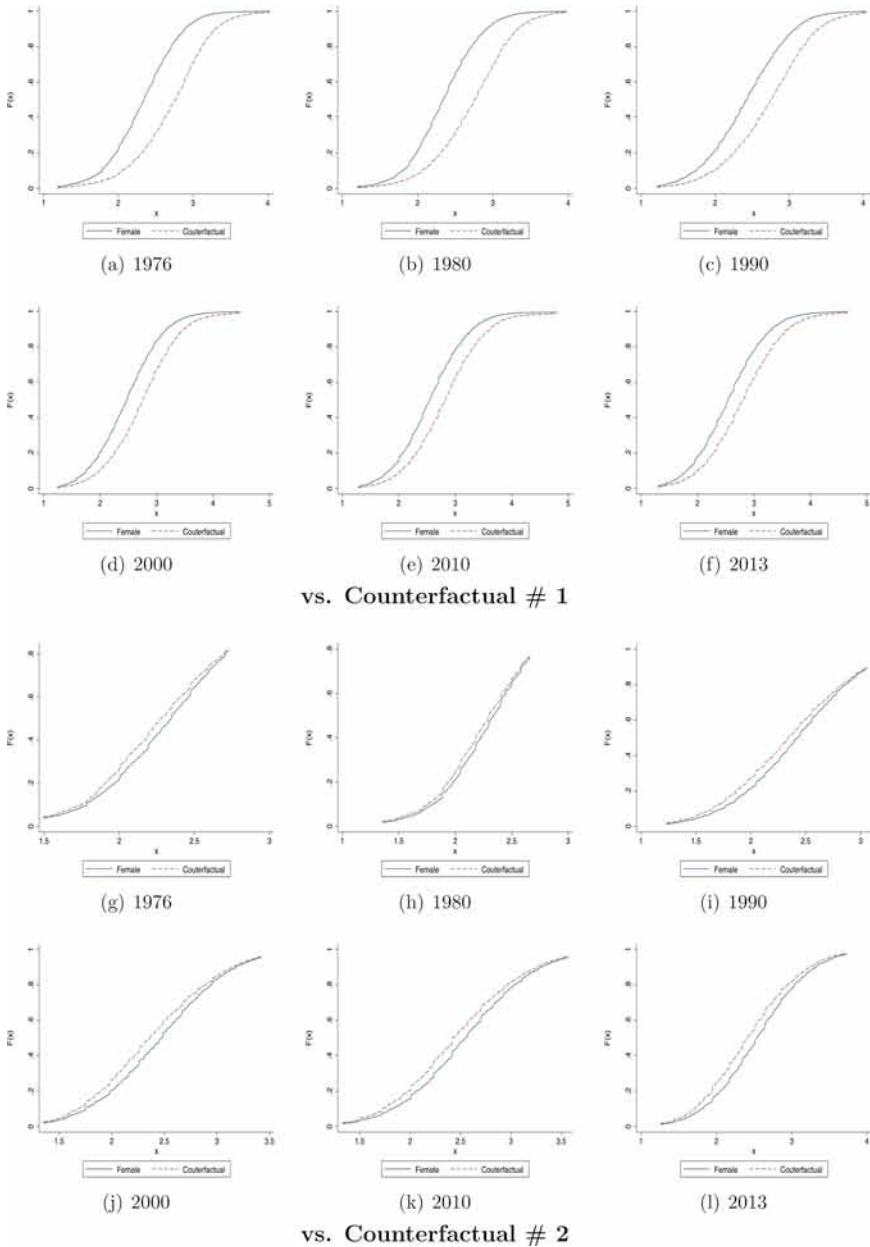
(k) 2010

(l) 2013

vs. Counterfactual # 2

Fig. 1. CDF of Female Versus Counterfactual Wage Distributions: Without Selection Correction.

Table 3. Female Wage Distribution Versus Counterfactual Distribution #2 (Without Selection Correction): Composition Effects.[a]

Year	srho	theil	mean	qte10	qte25	qte50	qte75	qte90
1976	0.207	0.415	−0.037	−0.037	−0.050	−0.069	−0.032	0.000
1977	0.160	0.321	−0.032	−0.030	−0.049	−0.055	0.000	0.000
1978	0.190	0.380	−0.032	−0.054	−0.070	−0.054	−0.017	0.004
1979	0.089	0.177	−0.024	−0.049	−0.045	−0.038	−0.007	0.011
1980	0.117	0.235	−0.027	−0.056	−0.036	−0.043	−0.010	0.008
1981	0.211	0.422	−0.037	−0.052	−0.079	−0.033	0.000	0.009
1982	0.172	0.343	−0.034	−0.077	−0.049	−0.044	−0.015	0.010
1983	0.197	0.394	−0.031	−0.066	−0.074	−0.017	0.000	0.022
1984	0.178	0.357	−0.039	−0.049	−0.067	−0.049	−0.007	0.000
1985	0.241	0.482	−0.048	−0.051	−0.100	−0.062	−0.027	0.003
1986	0.191	0.382	−0.049	−0.070	−0.059	−0.061	−0.011	−0.003
1987	0.241	0.501	−0.053	−0.095	−0.065	−0.065	−0.026	−0.004
1988	0.254	0.573	−0.056	−0.074	−0.099	−0.059	−0.015	0.000
1989	0.294	0.589	−0.059	−0.065	−0.107	−0.083	−0.031	0.000
1990	0.273	0.609	−0.063	−0.098	−0.087	−0.090	−0.041	−0.023
1991	0.345	0.690	−0.073	−0.075	−0.094	−0.086	−0.045	−0.017
1992	0.355	0.710	−0.074	−0.078	−0.090	−0.103	−0.053	−0.020
1993	0.306	0.613	−0.067	−0.069	−0.108	−0.105	−0.049	−0.010
1994	0.349	0.699	−0.074	−0.074	−0.102	−0.087	−0.065	0.000
1995	0.264	0.530	−0.066	−0.080	−0.090	−0.090	−0.034	−0.010
1996	0.330	0.660	−0.075	−0.049	−0.125	−0.096	−0.055	−0.028
1997	0.330	0.660	−0.073	−0.080	−0.100	−0.095	−0.051	−0.020
1998	0.333	0.667	−0.080	−0.085	−0.106	−0.118	−0.084	−0.033
1999	0.387	0.776	−0.086	−0.097	−0.102	−0.131	−0.079	−0.051
2000	0.347	0.695	−0.079	−0.082	−0.119	−0.118	−0.057	−0.030
2001	0.278	0.556	−0.070	−0.063	−0.118	−0.074	−0.064	−0.026
2002	0.296	0.593	−0.075	−0.095	−0.103	−0.079	−0.067	−0.031
2003	0.246	0.493	−0.066	−0.077	−0.098	−0.108	−0.038	−0.036
2004	0.330	0.660	−0.080	−0.074	−0.105	−0.103	−0.041	−0.051
2005	0.406	0.813	−0.090	−0.088	−0.105	−0.134	−0.049	−0.025
2006	0.395	0.791	−0.091	−0.094	−0.099	−0.118	−0.095	−0.036
2007	0.396	0.794	−0.095	−0.083	−0.100	−0.102	−0.085	−0.067
2008	0.392	0.785	−0.092	−0.099	−0.105	−0.090	−0.082	−0.053
2009	0.382	0.770	−0.092	−0.065	−0.129	−0.121	−0.105	−0.049
2010	0.339	0.679	−0.085	−0.105	−0.129	−0.133	−0.069	−0.038
2011	0.417	0.836	−0.096	−0.125	−0.143	−0.124	−0.075	−0.068
2012	0.449	0.898	−0.101	−0.094	−0.138	−0.115	−0.080	−0.061
2013	0.484	0.970	−0.107	−0.091	−0.147	−0.118	−0.080	−0.080

[a]Data Source: IPUMS CPS (http://cps.ipums.org/cps/). Columns (1) and (2) report the entropy gap measures ($\times 100$) at corresponding functionals of the distributions of log wages (measures the distance between the female and counterfactual wage). Columns (3)–(8) report conventional measures based on difference in parts of the wage distributions between the female wage distribution and the counterfactual distribution.

even less than one tenth of the structural effect counterpart. However, the magnitude has been increasing over time. The annualized rate of increase implied by the entropy measures is also about 5 percent. Moreover, unlike in the case of structural difference, we find that conventional measures trace out the pattern of our distributional measure well.

Turning to the SD results (Panel B of Table 2), we observe first-order dominance ranking of the female wage distribution and the counterfactual wage distribution (#2) in nearly all the cases. However, during the period 1976–1985, none of the results are statistically significant. The inability to rank order the earnings distributions between men and women in most cases is informative. This finding implies that any welfare conclusions concerning that women fare better or worse should they posse men's human capital characteristics in the labor market are not robust to changes in the particular welfare function being used, despite the fact that the observed differences in selected percentiles of the earnings distributions between men and women are in all the cases above. 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 promoting the policies aimed at improving women's human capital characteristics.

In more recent years, we observe statistically significant SD 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. This result is consistent with Goldin, Katz, and Kuziemko (2006): "by 1980, the college gender gap in enrollments had evaporated" and call this change a "homecoming" of American college women (to the parity observed in the early twentieth century)." As noted in Maasoumi and Wang (2019), this result is quite powerful, suggesting policies aimed at changing the human capital characteristics only, may not produce relative improvements for women.

5. RESULTS ADDRESSING SELECTION

To examine the impact of addressing the selection issue on the results above, we now turn to the inverse probability weighted estimators controlling for first-stage selection propensity scores. The comparison between the female wage distribution and the counterfactual distribution #1 is reported in Table 4. We find that addressing selection slightly impacts the counterfactual outcomes as a result of changes in wage structure. More important, the general pattern observed above continues to hold. Specifically, we again find that structure effects play an important role explaining the gender gap, but the implied size and trend vary with the conventional measures. When taking into account the entire distribution, our entropy measures suggest that the conventional measures appear to underestimate the rate of overall decrease of the structure effects in the society. Examining the actual SD test statistics in Panel A of Table 6, we again find statistically significant, first-order dominance relations in all cases. This result again indicates that any policymakers whose preferences are to increase women's wages would

Table 4. Female Wage Distribution Versus Counterfactual Distribution #1
(With Selection Correction): Structural Effects.[a]

Year	srho	theil	mean	qte10	qte25	qte50	qte75	qte90
1976	8.638	18.052	0.384	0.275	0.376	0.410	0.405	0.408
1977	8.291	17.258	0.375	0.268	0.372	0.402	0.407	0.405
1978	8.512	17.765	0.386	0.278	0.353	0.405	0.427	0.405
1979	8.080	16.803	0.375	0.222	0.350	0.422	0.437	0.427
1980	8.146	16.911	0.376	0.255	0.357	0.410	0.440	0.399
1981	8.008	16.750	0.371	0.258	0.331	0.411	0.429	0.414
1982	7.789	16.264	0.372	0.248	0.353	0.404	0.437	0.395
1983	6.638	13.737	0.350	0.246	0.309	0.397	0.386	0.392
1984	5.985	12.421	0.336	0.239	0.291	0.379	0.389	0.377
1985	5.828	12.081	0.337	0.250	0.288	0.369	0.377	0.401
1986	5.154	10.663	0.319	0.222	0.291	0.357	0.378	0.368
1987	4.768	9.755	0.317	0.238	0.300	0.353	0.342	0.343
1988	4.487	9.192	0.311	0.240	0.297	0.324	0.360	0.336
1989	4.622	9.884	0.316	0.248	0.298	0.346	0.335	0.350
1990	3.994	8.180	0.299	0.220	0.288	0.316	0.322	0.309
1991	3.692	7.523	0.286	0.222	0.282	0.327	0.307	0.316
1992	3.508	7.140	0.281	0.203	0.282	0.305	0.315	0.308
1993	3.201	6.517	0.265	0.182	0.252	0.288	0.303	0.303
1994	2.978	6.040	0.263	0.214	0.248	0.291	0.291	0.297
1995	2.872	5.815	0.265	0.190	0.254	0.288	0.288	0.291
1996	2.917	6.095	0.273	0.223	0.254	0.288	0.280	0.280
1997	3.081	6.279	0.282	0.222	0.288	0.301	0.289	0.309
1998	3.008	6.106	0.280	0.207	0.273	0.266	0.288	0.310
1999	3.296	6.758	0.289	0.233	0.292	0.288	0.288	0.293
2000	2.954	5.994	0.280	0.223	0.264	0.275	0.300	0.288
2001	3.026	6.162	0.288	0.232	0.257	0.293	0.292	0.339
2002	3.085	6.283	0.290	0.232	0.255	0.280	0.297	0.334
2003	2.640	5.382	0.266	0.204	0.220	0.251	0.288	0.326
2004	2.471	5.007	0.260	0.214	0.223	0.259	0.279	0.288
2005	2.662	5.413	0.273	0.237	0.247	0.259	0.301	0.328
2006	2.630	5.341	0.275	0.207	0.257	0.288	0.303	0.327
2007	2.415	4.883	0.269	0.236	0.257	0.255	0.272	0.306
2008	2.542	5.156	0.267	0.230	0.266	0.268	0.277	0.294
2009	2.463	4.988	0.266	0.223	0.256	0.262	0.280	0.330
2010	2.442	4.943	0.265	0.204	0.228	0.272	0.288	0.318
2011	2.345	4.742	0.260	0.177	0.223	0.251	0.302	0.289
2012	2.314	4.675	0.262	0.182	0.207	0.283	0.297	0.288
2013	2.147	4.336	0.255	0.192	0.203	0.280	0.288	0.262

[a]Data Source: IPUMS CPS (http://cps.ipums.org/cps/). Columns (1) and (2) report the entropy gap measures ($\times100$) at corresponding functionals of the distributions of log wages (measures the distance between the female and counterfactual wage). Columns (3)–(8) report conventional measures based on difference in parts of the wage distributions between the female wage distribution and the counterfactual distribution.

find favorable policies aimed at improving women's wage structure. The CDF comparisons of distributions for select years are provided in Fig. 2.

The comparison between the female wage distribution and the counterfactual distribution #1 is reported in Table 5. As we can see, the differences implied by all measures continue to be smaller, relative to the structure effects above. However,

(a) 1976   (b) 1980   (c) 1990

(d) 2000   (e) 2010   (f) 2013

vs. Counterfactual # 1

(g) 1976   (h) 1980   (i) 1990

(j) 2000   (k) 2010   (l) 2013
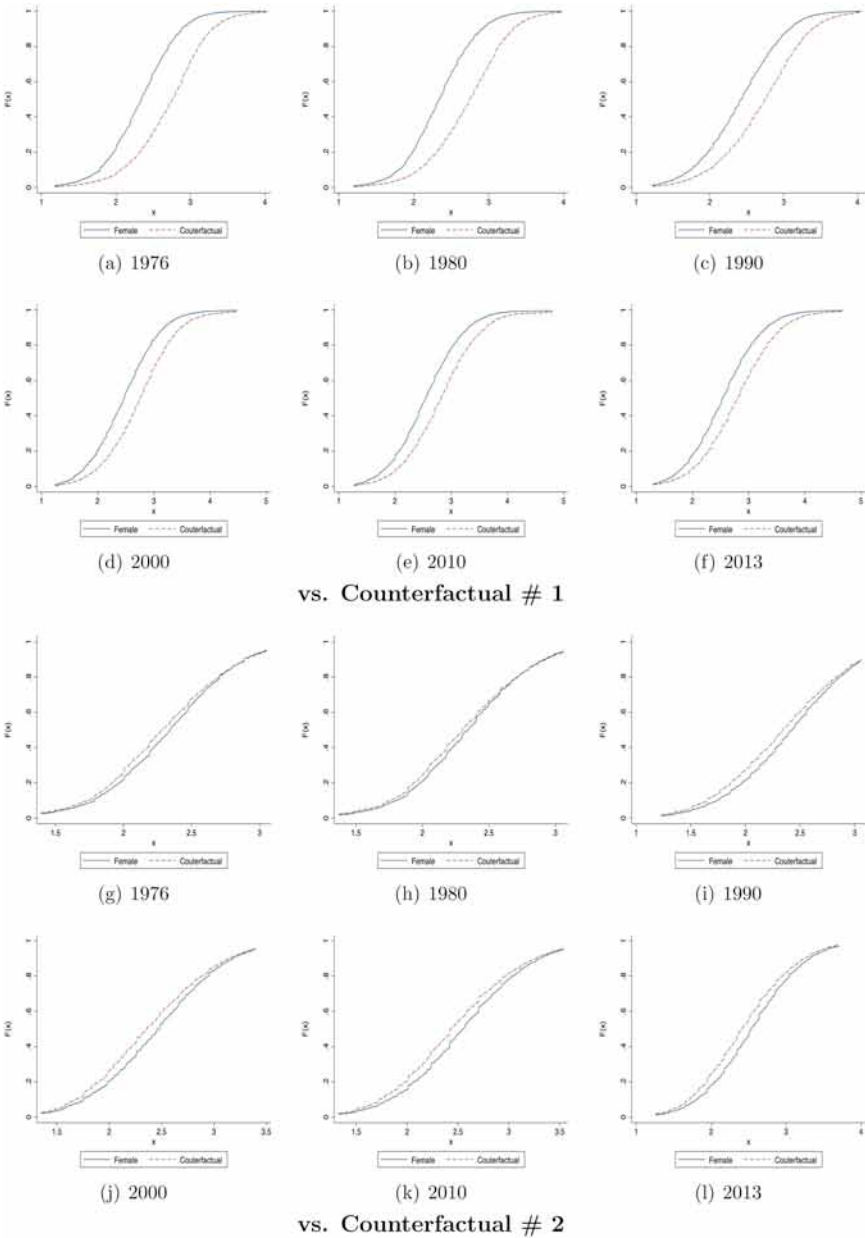
vs. Counterfactual # 2

Fig. 2. CDF of Female Versus Counterfactual wage Distributions: With Selection Correction.

Table 5.   Female Wage Distribution Versus Counterfactual Distribution #2
(Without Selection Correction): Composition Effects.[a]

Year	srho	theil	mean	qte10	qte25	qte50	qte75	qte90
1976	0.178	0.355	−0.034	−0.048	−0.041	−0.063	−0.031	0.000
1977	0.165	0.331	−0.032	−0.030	−0.049	−0.059	−0.001	0.000
1978	0.217	0.434	−0.035	−0.050	−0.070	−0.065	−0.018	0.004
1979	0.095	0.190	−0.024	−0.049	−0.045	−0.041	−0.007	0.013
1980	0.123	0.246	−0.027	−0.054	−0.037	−0.043	−0.010	0.013
1981	0.217	0.435	−0.038	−0.056	−0.080	−0.033	0.000	0.009
1982	0.214	0.429	−0.039	−0.080	−0.049	−0.049	−0.020	0.010
1983	0.200	0.400	−0.032	−0.070	−0.077	−0.017	0.000	0.022
1984	0.181	0.362	−0.038	−0.049	−0.070	−0.049	−0.006	0.000
1985	0.245	0.490	−0.047	−0.051	−0.100	−0.062	−0.025	0.008
1986	0.192	0.384	−0.048	−0.078	−0.060	−0.060	−0.011	0.000
1987	0.252	0.522	−0.053	−0.096	−0.065	−0.066	−0.025	0.000
1988	0.261	0.586	−0.057	−0.074	−0.099	−0.059	−0.017	0.000
1989	0.296	0.593	−0.059	−0.065	−0.107	−0.083	−0.033	0.000
1990	0.287	0.637	−0.064	−0.098	−0.087	−0.090	−0.041	−0.019
1991	0.333	0.668	−0.072	−0.071	−0.094	−0.082	−0.044	−0.011
1992	0.365	0.732	−0.074	−0.076	−0.090	−0.104	−0.053	−0.020
1993	0.307	0.615	−0.067	−0.069	−0.108	−0.105	−0.049	−0.010
1994	0.355	0.712	−0.075	−0.076	−0.106	−0.090	−0.065	0.000
1995	0.276	0.552	−0.066	−0.080	−0.090	−0.092	−0.034	−0.010
1996	0.346	0.692	−0.076	−0.045	−0.125	−0.098	−0.055	−0.028
1997	0.331	0.664	−0.074	−0.080	−0.098	−0.095	−0.051	−0.024
1998	0.335	0.670	−0.080	−0.086	−0.108	−0.118	−0.082	−0.029
1999	0.410	0.821	−0.088	−0.104	−0.105	−0.134	−0.079	−0.048
2000	0.348	0.696	−0.078	−0.077	−0.119	−0.118	−0.057	−0.029
2001	0.276	0.553	−0.069	−0.063	−0.118	−0.074	−0.062	−0.021
2002	0.292	0.585	−0.074	−0.095	−0.102	−0.079	−0.065	−0.031
2003	0.253	0.506	−0.066	−0.085	−0.105	−0.105	−0.034	−0.036
2004	0.329	0.660	−0.079	−0.074	−0.105	−0.103	−0.039	−0.049
2005	0.404	0.809	−0.089	−0.088	−0.105	−0.134	−0.049	−0.017
2006	0.396	0.793	−0.091	−0.093	−0.099	−0.118	−0.095	−0.036
2007	0.437	0.875	−0.098	−0.087	−0.102	−0.105	−0.085	−0.063
2008	0.386	0.774	−0.091	−0.096	−0.105	−0.087	−0.080	−0.053
2009	0.378	0.762	−0.091	−0.065	−0.129	−0.121	−0.105	−0.049
2010	0.339	0.679	−0.084	−0.105	−0.129	−0.133	−−0.066	−0.038
2011	0.417	0.835	−0.096	−0.125	−0.143	−0.124	−0.073	−0.068
2012	0.449	0.898	−0.100	−0.094	−0.140	−0.115	−0.080	−0.054
2013	0.486	0.973	−0.107	−0.090	−0.147	−0.118	−0.077	−0.078

[a]Data Source: IPUMS CPS (http://cps.ipums.org/cps/). Columns (1) and (2) report the entropy gap measures ($\times 100$) at corresponding functionals of the distributions of log wages (measures the distance between the female and counterfactual wage). Columns (3)–(8) report conventional measures based on difference in parts of the wage distributions between the female wage distribution and the counterfactual distribution.

addressing selection does impact the estimates to a much greater extent. For example, in 1976, the difference at the 10th percentile when addressing selection is about 30 percent larger than the estimate without addressing selection. This result is similarly reflected in our entropy measures. Nevertheless, the long-run trend observed above continues to hold. We again find that the role of human capital characteristics in affecting women's wages has continued to increase.

Turning to our SD results in Panel B of Table 6, we also find that addressing selection impacts our analysis. We fail to observe first-order dominance in most years during the period 1976–1983. However, we do observe a few instances of second-order dominance relations. This result means that even though there are losers and winners, the losers are mostly concentrated in the upper tail. As a result, only individuals with social welfare function increasing in wage and averse to inequality would conclude there exists a welfare improvement for women from changing the human capital characteristics. Nevertheless, these results are not statistically significant. When examining later years, we again find significant dominance relations, indicating improving women's human capital characteristics does not necessarily improve their wages.

6. CONCLUSIONS

In this chapter, we present a set of complementary tools that move beyond the simple moment-based comparison of the earnings distribution and the counterfactual distributions. In particular, we discuss entropy measures based on the distance between two *whole* earnings distributions, instead of their specific parts. We also discuss tests based on stochastic dominance to allow for robust welfare comparisons of the female earnings distribution and the counterfactual distribution. Building on recent advances in the treatment effects literature to identify the counterfactual distributions and using the CPS data 1976–2013, we illustrate this framework in the context of the gender gap in the United States. We reach two main conclusions. First, we find that regardless of measures used, structure effects are much larger than composition effects. But the importance of structure effects has decreased over time, while that of composition has increased. Moreover, traditional moment-based measures severely underestimate the declining trend of the structure effects in the United States Second, we find first-order stochastic dominance in all cases when comparing the female distribution and the counterfactual distribution when women are endowed with men's wage structure. This result is powerful, suggesting policies aimed at increasing women's pay equity could potentially improve women's welfare uniformly. In contrast, in early years, we fail to find any statistically significant dominance relations when comparing the female distribution to the counterfactual distribution when women possess the same distribution of human capital characteristics as men do. In later years, we do find dominance relations, but the results suggest that women's human capital characteristics are not necessarily inferior to that of men's, and thus policies aimed at changing the human capital characteristics only, may not produce relative improvements for women. Finally, addressing selection impacts primarily the counterfactual distribution when changing the distribution of women's human

Table 6. Stochastic Dominance Results with Selection Correction.

Year	Observed Ranking	d	$Pr[d \leq 0]$	s	$Pr[s \leq 0]$	Observed Ranking	d	$Pr[d \leq 0]$	s	$Pr[s \leq 0]$
	Panel A: vs Counter-factual #1					Panel B: vs Counter-factual #2				
1976	FSD	−0.640	1.000	−0.640	1.000	SSD	0.681	0.344	−0.647	0.997
1977	FSD	−0.727	1.000	−0.727	1.000	SSD	0.714	0.408	−0.752	1.000
1978	FSD	−0.705	1.000	−0.705	1.000	SSD	0.833	0.304	−0.710	1.000
1979	FSD	−0.701	1.000	−0.718	1.000	SSD	0.739	0.314	−0.753	1.000
1980	FSD	−0.782	1.000	−0.808	1.000	SSD	0.848	0.324	−0.834	1.000
1981	FSD	−0.780	1.000	−0.780	1.000	FSD	−0.788	0.769	−1.037	1.000
1982	FSD	−0.736	1.000	−0.763	1.000	FSD	−0.764	0.585	−0.764	1.000
1983	FSD	−0.732	1.000	−0.910	1.000	SSD	1.164	0.074	−0.734	1.000
1984	FSD	−0.724	1.000	−0.726	1.000	FSD	−0.723	0.672	−0.723	1.000
1985	FSD	−0.744	1.000	−0.757	1.000	FSD	−0.721	0.435	−0.895	1.000
1986	FSD	−0.792	1.000	−0.930	1.000	FSD	−0.760	0.933	−0.774	1.000
1987	FSD	−0.749	1.000	−0.751	1.000	FSD	−0.747	0.930	−0.794	1.000
1988	FSD	−0.783	1.000	−1.157	1.000	FSD	−0.841	1.000	−0.842	1.000
1989	FSD	−0.827	1.000	−0.921	1.000	FSD	−0.781	0.856	−0.814	1.000
1990	FSD	−0.950	1.000	−0.950	1.000	FSD	−0.834	1.000	−0.834	1.000
1991	FSD	−0.914	1.000	−0.916	1.000	FSD	−0.817	0.993	−0.961	1.000
1992	FSD	−0.830	1.000	−0.835	1.000	FSD	−0.886	1.000	−1.205	1.000
1993	FSD	−0.783	1.000	−0.879	1.000	FSD	−0.804	1.000	−0.835	1.000
1994	FSD	−0.794	1.000	−0.876	1.000	FSD	−0.900	1.000	−0.966	1.000
1995	FSD	−0.844	1.000	−0.844	1.000	FSD	−0.821	1.000	−0.821	1.000
1996	FSD	−0.821	1.000	−0.836	1.000	FSD	−0.729	0.997	−0.738	0.997
1997	FSD	−0.747	1.000	−0.747	1.000	FSD	−0.744	1.000	−0.758	1.000
1998	FSD	−0.722	1.000	−0.807	1.000	FSD	−0.746	1.000	−0.747	1.000
1999	FSD	−0.827	1.000	−0.960	1.000	FSD	−0.776	1.000	−0.815	1.000
2000	FSD	−0.788	1.000	−0.788	1.000	FSD	−0.779	1.000	−0.855	1.000
2001	FSD	−1.061	1.000	−1.063	1.000	FSD	−0.903	1.000	−1.162	1.000
2002	FSD	−0.995	1.000	−0.995	1.000	FSD	−0.975	1.000	−1.203	1.000
2003	FSD	−0.966	1.000	−1.045	1.000	FSD	−0.958	1.000	−0.958	1.000
2004	FSD	−1.027	1.000	−1.030	1.000	FSD	−0.981	1.000	−1.182	1.000
2005	FSD	−0.993	1.000	−0.993	1.000	FSD	−0.952	0.997	−0.952	1.000
2006	FSD	−1.003	1.000	−1.003	1.000	FSD	−0.940	1.000	−1.006	1.000
2007	FSD	−0.959	1.000	−0.959	1.000	FSD	−0.918	1.000	−0.942	1.000
2008	FSD	−0.963	1.000	−0.963	1.000	FSD	−0.955	1.000	−1.414	1.000
2009	FSD	−1.052	1.000	−1.089	1.000	FSD	−0.947	1.000	−1.051	1.000
2010	FSD	−0.922	1.000	−0.922	1.000	FSD	−0.925	1.000	−0.935	1.000
2011	FSD	−0.918	1.000	−1.180	1.000	FSD	−0.916	1.000	−0.916	1.000
2012	FSD	−0.937	1.000	−1.014	1.000	FSD	−0.899	1.000	−0.964	1.000
2013	FSD	−0.905	1.000	−0.952	1.000	FSD	−0.905	1.000	−1.001	1.000 '

capital characteristics to that of men's, but not the alternative counterfactual distribution when changing the wage structure. Despite the changes of the estimate, the general patterns remain the same.

NOTES

1. 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 the corresponding distributions.

2. For the proofs of these equivalent definitions, see, e.g., Chapter 1 of Whang (2019).

3. In the Tables in online appendix, we also report $d_1 = \sup[F_{g|T=t}(y) - F_{Y(1)|D=1,T=t}(y)]$ and $d_2 = \sup[F_{Y(1)|D=1,T=t}(y) - F_{g|T=t}(y)]$. These two numbers will help us draw the conclusion of the direction of dominance. For example, if $d \leq 0$ and $d \leq 0$, then the former FSD the latter distribution. s_1, s_2 are similarly defined.

4. Note that $\tilde{X}$ is assumed to be independent of both error terms throughout the chapter.

5. Our *Assumption 4.2* is similar to Arellano and Bonhomme's Assumption 1.A1, except that we have a treatment here that is included in the assumption. Our assumption does not preclude the possibility that the dependence between unobservables can also depend on the covariates, X.

6. Note that in this process, the first-stage estimation of propensity scores is parametric while the construction of weights and counterfactual distributions are nonparametric. One can employ either parametric or nonparametric binary models.

7.

$$\widehat{\omega_1(D_i)} = \frac{D_i}{\hat{p}} / \sum_{i=1}^{N}\left[\frac{D_i}{\hat{p}}\right] \quad \text{and} \quad \widehat{\omega_0(D_i)} = \frac{1-D_i}{1-\hat{p}} / \sum_{i=1}^{N}\left[\frac{1-D_i}{1-\hat{p}}\right]$$

where $\hat{p} = \frac{1}{N}\sum_{i=1}^{N_s} D_i$.

$$\widehat{\omega_c(D_i)} = \frac{1-D_i}{\hat{p}}\frac{\pi\left(X_i, \widehat{p(W_i)}\right)}{1-\pi\left(X_i, \widehat{p(W_i)}\right)} / \sum_{i=1}^{N}\left[\frac{1-D_i}{\hat{p}}\frac{\pi\left(X_i, \widehat{p(W_i)}\right)}{1-\pi\left(X_i, \widehat{p(W_i)}\right)}\right]$$

8. Also noted in Machado (2012), the number of children is used as an explanatory variable in the shadow price function in Heckman (1974), "one of the seminal works on female selection," and IV in the participation equation in Heckman (1980). The number of young children may affect women's reservation wages and their labor supply decisions because it could affect "the value of leisure" for women (Keane, Wolpin, & Todd, 2011) and child-rearing is time consuming and costly.

ACKNOWLEDGMENTS

REFERENCES

Blau, F. D., & Kahn, L. M. (1997). Swimming upstream: Trends in the gender wage differential in the 1980s. *Journal of Labor Economics*, *15*, 1–42.

Blau, F. D., & Kahn, L. M. (2006). The U.S. gender pay gap in the 1990s: Slowing convergence. *Industrial and Labor Relations Review*, *60*, 45–66.

Bollinger, C. R., & Hirsch, B. T. (2013). Is earnings nonresponse ignorable? *The Review of Economics and Statistics*, *95*, 407–416.

Eren, O., & Ozbeklik, S. (2014). Who benefits from job corps? A distributional analysis of an active labor market program. *Journal of Applied Econometrics*, *29*, 586–611.

Firpo, S., Fortin, N., & Lemieux, T. (2007). *Decomposing wage distributions using recentered influence function regressions*. Unpublished Manuscript.

Flood, S., King, M., Ruggles, S., & Robert Warren, J. (2017). *Integrated Public use microdata series, current population survey: Version 5.0. [dataset]*. Technical report, University of Minnesota.

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition methods in economics. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 4, pp. 1–102). Amsterdam, Netherlands: Elsevier.

Goldin, C., Katz, L. F., & Kuziemko, I. (2006). The homecoming of American college women: The reversal of the college gender gap. *Journal of Economic Perspectives*, *20*, 133–156.

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

Gunderson, M., & Riddell, W. C. (1992). Comparable worth: Canada's experience. *Contemporary Policy Issues*, *10*, 85–94.

Hartmann, H. I., & Aaronson, S. (1994). Pay equity and women's wage increases: Success in the States, a model for the nation. *Duke Journal of Gender Law and Policy*, *1*, 69–87.

Heckman, J. J. (1974). Shadow price, market wages, and labor supply. *Econometrica*, *42*, 679–694.

Heckman, J. J. (1980). Sample selection bias as a specification error with an application to the estimation of labor supply functions. In *Female labor supply: Theory and estimation*. Princeton, NJ: Princeton University Press.

Huber, M. (2014). Treatment evaluation in the presence of sample selection. *Econometric Reviews*, *33*, 869–905.

Huber, M., & Mellace, G. (2014). Testing exclusion restrictions and additive separability in sample selection models. *Empirical Economics*, 2011–2045.

Keane, M. P., Wolpin, K. I., & Todd, P. (2011). The structural estimation of behavioral models: discrete choice dynamic programming methods and applications. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 4, pp. 331–461). Amsterdam, Netherlands: Elsevier.

Linton, O., Maasoumi, E., & Whang, Y. J. (2005). Consistent testing for stochastic dominance: A subsampling approach. *Review of Economic Studies*, *72*, 735–765.

Maasoumi, E., & Heshmati, A. (2000). Stochastic dominance amongst Swedish income distributions. *Econometric Reviews*, *19*, 287–320.

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

Maasoumi, E., & Wang, L. (2017). What can we learn about the racial gap in the presence of sample selection? *Journal of Econometrics*, *199*, 117–130.

Maasoumi, E., & Wang, L. (2019). The gender gap between earnings distributions. *Journal of Political Economy*, *127*, 2438–2504.

Machado, C. (2012). *Selection, heterogeneity and the gender wage gap*. IZA Discussion Papers 7005.

Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory*. Oxford: Oxford University Press.

Millimet, D. L., & Wang, L. (2006). A distributional analysis of the gender earnings gap in Urban China. *The B.E. Journal of Economic Analysis & Policy (Contributions)*, *5*, Article 5.

Mulligan, C. B., & Rubinstein, Y. (2008). Selection, investment, and women's relative wages over time. *Quarterly Journal of Economics*, *123*, 1061–1110.

Polachek, S. W. (2006). How the life-cycle human-capital model explains why the gender wage gap narrowed. In *The declining significance of gender?* (pp. 102–124). New York, NY: Russell Sage Foundation.

Skaug, H. J., & Tjostheim, D. (1996). Measures of distance between densities with application to testing for serial independence. In P. M. Robinson & M. Rosenblatt (Eds.), *Time series analysis in memory of E.J. Hannan* (Vol. 2, pp. 363–377). New York, NY: Springer.

Waldfogel, J., & Mayer, S. E. (2000). Gender differences in the low-wage labor market. In *Finding jobs: Work and welfare reform* (pp. 193–232). New York, NY: Russell Sage Foundation.

Whang, Y.-J. (2019). *Econometric analysis of stochastic dominance: Concepts, methods, tools, and applications.* Cambridge: Cambridge University Press.