Decomposing Wage Changes in the United States

Iván Fernández-Val* Franco Peracchi† Aico van Vuuren‡
Francis Vella§

May 22, 2018

Abstract

We analyze the patterns and determinants of hourly real wage changes in the United States for the period 1976 to 2016 at various quantiles of the wage distribution while accounting for selection bias from the employment decision. Male wages at the median and below have decreased despite large increases in both the skill premia and the average level of educational attainment. Female wage growth at the lower quantiles is small. Wages at the upper quantiles of the distribution have increased for both genders. We find that composition effects have increased wages at all quantiles but that the movements in the structural effects have generally determined the overall pattern of wages. There is only evidence of changes in the selection effects for the lower quantiles of the female wage distribution. We find that these components have combined to produce a substantial increase in wage inequality. The increased participation of females has further exacerbated female wage inequality.

Keywords: Wage inequality, wage decompositions, sample selection
JEL-codes: C14, I24, J00

---

*Boston University
†Georgetown University
‡Gothenburg University
§Georgetown University

1 Introduction

Substantial and increasing income inequality is associated with a variety of economic and social concerns. Accordingly, the dramatic increase in earnings inequality in the United States has been, and continues to be, an intensely studied phenomenon (see, for example, Katz and Murphy 1992, Murphy and Welch, 1992, Juhn, et al., 1993, Katz and Autor, 1999, Lee, 1999, Lemieux, 2006, Autor, et al., 2008, Acemoglu and Autor, 2011, and Murphy and Topel, 2016). While contributions to this literature employ a variety of measures of earnings and inequality and utilize different data sets, there is a general agreement that earnings inequality has greatly increased since the early 1980s. As an individual's earnings reflects both the number of hours worked and the average hourly wage rate received for those hours (Gottschalk and Danzinger, 2005), understanding the determinants of each of these factors is important for uncovering the sources of earnings inequality. This paper examines the evolution and determinants of changes in hourly wage rates in the United States for the period 1976 to 2016.

Figure 1 presents the profiles of selected quantiles of male and female real hourly wage rates using data from the U.S. Current Population Survey, or March CPS, for the survey years 1976 to 2016. They are based on individuals aged 24 to 65 years at the survey date who worked a positive number of hours the previous year and are neither in the Armed Forces nor self-employed.[1] The profiles differ greatly by gender and quantile. For example, consider the median male wage rate. Despite a slight increase between 1987 and 1990, the overall trend between 1976 and 1994 was negative, reaching a minimum of 17.6 percentage points below its initial level. It rebounds between 1994 and 2002, but falls between 2007 and 2013. The decline between 1976 and 2016 is 13.6 percentage points. The female median wage, despite occasional dips, increases by nearly 25 percentage points over the sample period. While the decrease in the male median wage is concerning, the male wage profile at the 25th percentile is alarming. This profile shows the same cyclical behavior as the median but a greater decline, especially between 1976 and 1994, resulting in a 2016 wage level that is 18.2 percentage points below its 1976 level. In contrast, the

[1] Section 2 provides a detailed discussion of the data employed in our analysis and our sample selection process.

female wage at the 25th percentile increases by about 17 percentage points. The time trend of the profile at the 10th percentile is similar to those of the 25th and 50th percentiles.

Descriptions of wage inequality involve contrasts between the lower and upper parts of the wage distribution and thus we also examine profiles of the higher quantiles. There is a small increase in the male real wage at the 75th percentile over the whole period. However, until the late 1990's it was below its 1976 level before experiencing a notable increase in the last 15 years. The profile at the 90th percentile resembles the 75th percentile, although the periods of growth have produced larger increases. For females, there has been a strong and steady growth at both the 75th and 90th percentiles since 1980 with an increasing gap between each of them and the median wage.

Figure 2 reports the time series behavior of the 90th/10th percentile ratios and reveals a widening wage gap for both males and females with increases over the sample period of 54.6 and 37.9 percent respectively. Rather than immediately focusing on the factors generating these ratios, we first direct our attention on identifying the determinants of the observed wage changes. The cyclicality of the profiles in Figure 1 suggests that each responds to business cycle forces, although the strength of this relationship varies by quantile. Perhaps more notably, the behavior of the profiles at the upper and lower quantiles suggest that workers in these associated labor markets are undergoing different experiences. Our objective is to understand what is generating these trends.

Previous empirical work has identified some possible determinants of the growth in wage inequality. First, the labor economics literature (see for example, Juhn, et al., 1993, Katz and Autor, 1999, Welch, 2000, Autor, et al., 2008, Acemoglu and Autor, 2011, and Murphy and Topel, 2016) has highlighted the role of increasing skill premia and especially the increasing returns to higher education. The effect by which the prices of individual's characteristics contribute to wages is known as the "structural effect". This also captures factors such as declining real minimum wages and the decrease in the union premium (see, for example, DiNardo, Fortin and Lemieux, 1996 and Lee, 1999), which are particularly important for the lower part of the wage distribution. Krueger and Posner (2018) argue that employers have also reduced the bargaining power of relatively low paid workers through the inclusion of non-compete clauses in employment contracts. In contrast to some earlier work, we do not attempt to disentangle the different factors in this structural effect. Rather, we combine the various contributions into one total effect. Second, the nature of the workforce has changed over the past 41 years, suggesting that changes in workers' characteristics have also contributed to wage movements. This is primarily reflected by the increases in educational attainment, decreases in unionization rates, and the changes in the age structure of the workforce caused by the aging of the Baby Boom cohorts. The large increase in female labor force participation may have also produced changes in the composition of the labor force composition. The contribution to changes in the distribution of hourly wages hourly wages attributable to these observed characteristics is known as the "composition effect".

Earlier papers (see, for example, Angrist, Chernozhukov and Fernàndez-Val, 2006, and Chernozhukov, Fernández-Val and Melly, 2013) have estimated structural and composition effects in general forms and under general conditions. However, they typically have not incorporated the potential selection bias resulting from the non randomness of the employment outcome (see Heckman, 1974, 1979). Thus, they have not allowed for the changing nature, in terms of unobservables, of the workforce. This "selection effect", is potentially important as the movements in the participation rates and in the average number of annual hours worked of both males and females, shown in Figure 3, suggest that the workforce may have changed along dimensions that are not directly observable. Figure 4 shows the fraction of wage earners who work either full-time or full-year. For males, it fluctuates cyclically between 80 and 90 percent. For females, despite the positive labor market trends during our sample period, it remains below 75 percent in 2016. Moreover, just as the returns to observed characteristics may evolve over time, and differ across different parts of the wage distribution, the return to unobservables may also vary. Mulligan and Rubinstein (2008) find that selection by females into full-time employment played an important role in explaining the variation in inequality. They find that the selected sample of working females became increasingly more productive in terms of unobservables during the 70's, 80's and 90's of the twentieth century and this explains the reduction in the gender wage gap in that period. Moreover they argue that their evidence is consistent with the observed increase in wage inequality for females. That is, the increased productivity level of the selected sample reflects the greater incentive for the most (least) productive women to enter (leave) the labor market. We do not find selection effects of the same magnitude as Mulligan and Rubinstein (2008). However, as discussed below, our definition of the selection effect is not directly comparable to theirs.

While accounting for selection in the estimation of the conditional mean within a separable model, as is done in Mulligan and Rubinstein (2008), is straightforward, it is more challenging when estimating the determinants of wages at different quantiles. To do so requires a nonseparable model. Arellano and Bonhomme (2017) adopt a copula approach to modeling the wage distribution with a binary selection rule capturing the work decision. Fernández-Val, Van Vuuren and Vella (2018), hereafter FVV, propose an estimation strategy for nonseparable models with a censored selection rule. They employ distribution regression methods and account for sample selection via an appropriately constructed control function. We employ the FVV approach to nonparametrically estimate the relationship between the individual's wages and their characteristics while accounting for selection. Although this requires more information on the selection variable this is provided in our data set in the form of the number of annual working hours.

The following section discusses the data. Section 3 provides a discussion of our empirical model, the objects of interest and the estimation procedure. Section 4 presents the empirical results. Section 5 provides some additional discussion of the empirical results. Concluding comments are offered in Section 6.

2 Data

We employ micro-level data from the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS), or March CPS, for each of the 41 survey years from 1976 to 2016 which report annual earnings for the calendar years from 1975 to 2015.[2] We begin with the 1976 survey as this is the first period

[2] We downloaded the data from the IPUM-CPS website maintained by the Minnesota Population Center at the University of Minnesota (Flood et al. 2015), which provides standardized data from 1963 to 2016.

for which information on weeks worked and usual hours of work per week last year are available. To ensure that most individuals have completed their educational process and to avoid issues related to retirement we restrict attention to those aged 24–65 years in the survey year. This produces an overall sample of 1,794,466 males and 1,946,957 females, with an average annual sample size of 43,767 males and 47,487 females. There are large fluctuations in annual sample sizes, ranging from a minimum of 30,767 males and 33,924 females in 1976 to a maximum of 55,039 males and 59,622 females in 2001.

Annual hours worked last year are defined as the product of weeks worked and usual weekly hours of work last year. Most of those reporting zero hours also respond that they are not in the labor force (i.e., they report themselves as doing housework, unable to work, at school, retired or other) in the week of the March survey. We define hourly wages as the ratio of reported annual earnings and annual hours worked last year. Hourly wages are unavailable for those not in the labor force. This definition of hourly wages creates an issue for the Armed Forces, the self-employed, and unpaid family workers as their annual earnings or annual hours tend to be poorly measured. To bypass this issue, we exclude the Armed Forces, the self-employed, and unpaid family workers from our sample and focus attention on civilian dependent employees with positive hourly wages and people out of the labor force last year. This restricted sample contains 1,551,796 males and 1,831,220 females (respectively 86.5 percent and 94.1 percent of the original sample of people aged 24–65), with average annual sample sizes of 37,849 males and 44,664 females. The subsample of civilian dependent employees with positive hourly wages contains 1,346,918 males and 1,276,125 females, with an average annual sample size of 32,852 males and 31,125 females.

The March CPS differs from the commonly employed Outgoing Rotation Groups of the CPS, or ORG CPS, which contains information on hourly wages in the survey week for those who are paid by the hour and on weekly earnings from the primary job during the survey week for those who are not paid by the hour. Lemieux (2006) and Autor, Katz and Kearney (2008) argue that the ORG CPS data are preferable because they provide a point-in-time wage measure and because workers paid by the hour (more than half of the U.S. workforce) may recall their hourly wages better.

However, there is no clear evidence regarding differences in the relative reporting accuracy of hourly wages, weekly earnings and annual earnings. In addition, many workers paid by the hour also work overtime, so their effective hourly wage depends on the importance of overtime work and the wage differential between straight time and overtime. Furthermore, the failure of the March CPS to provide a point-in-time wage measure may actually be an advantage as it smooths out intra-annual variations in hourly wages.

There are concerns with using earnings data from either of the CPS files. First, defining hourly wages as the ratio of earnings (annual or weekly) to hours worked (annual or weekly) may induce a "division bias" (see, e.g., Borjas 1980). Second, CPS earnings data are subject to measurement issues, including top-coding of earnings (see Larrimore et al. 2008), mass points at zero and at values corresponding to the legislated minimum wage (DiNardo, Fortin and Lemieux 1996), item non-response, especially in the tails of the earnings distribution, which has been rising sharply over time (see, e.g., Meyer, Mok and Sullivan 2015), earnings response from proxies, and earnings imputation procedures by the Census Bureau (see Lillard, Smith and Welch 1986, and Bollinger et al. 2018). Our use of distribution and quantile regression methods mitigates these first two concerns. Since there is no consensus on how to best address the remaining three, we retain proxy responses and imputed earnings.

To explain the variation in hourly wage rate and annual hours of work, we employ the following conditioning variables; the individual's age and categorical variables for highest educational attainment (less than high-school, high-school graduate, some college, and college or more), race (white and non white), region of residence (North-East, South, Central, and West), and marital status (married with spouse present, and not married or spouse not present). We also employ household composition variables, including the number of members and the number of children in the household, and indicators for the presence of children under 5 years of age and the presence of other unrelated individuals in the household.

3 Model and Objects of Interest

Our primary objective is to decompose the movements over time in individual hourly wage rates into composition, structural and selection effects. Moreover, we perform these decompositions at various points in the wage distribution. We do this by estimating a model for the observed distribution of wages while accounting for selection and then estimating the objects of interest which capture these components. We also pay particular attention to the impact of education on wages.

3.1 Model

We consider a model based on the Heckman (1979) sample selection model, where the censoring rule generating the selection process incorporates that we observe the annual number of hours worked, and not just the binary work/not work rule. Our model has the general form:

$$W = g(X, \varepsilon), \text{ if } H > 0, \tag{1}$$

$$H = \max\{h(X, Z, \eta), 0\}, \tag{2}$$

where W denotes the log hourly wage rate, H denotes annual hours worked, and X and Z are vectors of observable explanatory variables. The functions g and h are unknown, and ε and η are respectively a vector and a scalar of potentially dependent unobservables with a strictly increasing distribution function. We also assume that ε and η are independent from X and Z in the total population and that the function h is strictly increasing in its last argument. The model is a nonparametric and nonseparable representation of the tobit type-3 model previously estimated under a variety of parametric and semi-parametric settings (see, for example, Amemiya, 1984, Vella, 1993, Honoré et al., 1997, Chen, 1997, and Lee and Vella, 2006). The most general treatment, and best suited to our objectives, is provided by FVV.

Due to the potential dependence between ε and η, the assumption that X and Z are independent of ε in the total population does not exclude their dependence in the selected sample. However, FVV show that ε is independent of X and Z conditional on $V \equiv F_{H|X,Z}(H|X, Z)$ and $H > 0$, which implies that $F_{H|X,Z}(H|X, Z)$

is an appropriate control function in this setting.[3] The intuition behind this result is that one can show that $V = F_\eta(\eta)$ and thus V can be interpreted as an index for the hours equation error term. This control function can be estimated via a distribution regression of H on X and Z.

We denote the support of the selected sample by $\mathcal{XV}$. This set consist of the potential combinations of observed characteristics X and values of the control function V for the individuals with a positive number of working hours. We define the set $\mathcal{V}(x)$ as the support of V conditional on a particular level of $X = x$ in the selected sample.

3.2 Local Effects

As previous studies partially attribute the increase in inequality to the changing skill premium, we focus on how the returns to education have evolved in the presence of changing participation rates and hours of work. We estimate what FVV define as "local objects". These are functionals of the distribution of W conditional on the values of the control function V and a given set of observed characteristics X. Local objects are defined for the total population and are not restricted to the selected sample. Two important local objects correspond to the mean and the distribution of W. These are the local average structural function (LASF) and the local structural distribution function (LSDF).

The LASF is defined as:

$$\mu(x, v) = \mathbb{E}[g(x, \varepsilon) \mid V = v],$$

i.e. the mean of W if all observations with a control function equal to v had characteristics x. The LSDF is defined as:

$$G(w, x, v) = F_{g(x,\varepsilon)|V}(w|v) = \mathbb{E}[\mathbf{1}\{g(x, \varepsilon) \leq w\} \mid V = v],$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. This is the distribution of W if all observations with a control function equal to v had characteristics x. The LASF

[3] This result is closely related to that of Imbens and Newey (2009).

and LSDF are identified provided the combination of x and v considered belongs to the support of the selected sample, that is $(x, v) \in \mathcal{XV}$. This also implies that for a given x, these objects are identified on the observed set $\mathcal{V}(x)$.[4] The LASF can be estimated using a flexible regression of W on X and V and the LDSF can be estimated using flexible distribution regression methods. This is explained in the appendix.

From the LDSF we derive the local quantile structural function (LQSF) as:

$$q(\tau, x, v) = \inf\{w \in \mathbb{R} : G(w, x, v) \geq \tau\}.$$

The LQSF is the left-inverse function of $w \mapsto G(w, x, v)$ and corresponds to the quantiles of W for a given set of observed characteristics X conditional on a particular value of V.

3.3 Global Effects and Wage Decompositions

FVV derive global counterparts of the local effects by integrating over the control variable. The general form of a global effect at $x \in \mathcal{X}$ is:

$$\theta(x|x_0) = \int_{\mathcal{V}(x_0)} \theta(x, v) dF_{V|X,H>0}(v|x_0), \tag{3}$$

where $\theta(x, v)$ can be any of the local objects defined above, $F_{V|X,H>0}(v|x_0)$ is the distribution function of V conditional on the outcome $X = x_0$, and $\mathcal{V}(x_0)$ is the corresponding support. Since the distribution function $F_{V|X,H>0}(v|x_0)$ is identified, identification of $\theta(x|x_0)$ requires the identification of $\theta(x, v)$ over $\mathcal{V}(x_0)$. From above we know that, for any given value of x, local objects are identified on the set of $\mathcal{V}(x)$. Hence, identification of a global object requires $\mathcal{V}(x_0) \subset \mathcal{V}(x)$. That is, individuals with a positive probability of being in the selected sample with observed characteristics x_0 also have a positive probability of being selected if their observed characteristics are equal to x. The appropriateness of this restriction depends on the situation. For example, it seems reasonable to assume that all low educated

[4]FVV provide the proofs of these and related identification results. The proofs are constructive in that they show that these objects are identical to the objects of W conditional on $X = x$, $V = v$ and conditional on the selected sample.

workers with a positive probability of working will also have a positive probability of working when assigned a higher level of education.

An interesting global object is the average structural function (ASF) conditional on $X = x_0$ in the selected population:

$$\mu^s_{x_0}(x) = \mathbb{E}[g(x, \varepsilon) \mid H > 0, X = x_0],$$

where the superscript s indicates that this object is for the selected population. This represents the average wage in the selected population of individuals with X equal to x_0 when their observed characteristics are set equal to x. The average treatment effect of changing X from x_0 to x_1 in the selected population is therefore:

$$\mu^s_{x_0}(x_1) - \mu^s_{x_0}(x_0). \tag{4}$$

Similarly, one can consider the structural distribution function (SDF) in the selected population as in Newey (2007). That is:

$$G^s_{x_0}(w, x) = \mathbb{E}[\mathbf{1}\{g(x, \varepsilon) \leq w\} \mid X = x_0, H > 0], \tag{5}$$

which gives the distribution of the potential outcome $g(x, \varepsilon)$ at w in the selected population among individuals with characteristics x_0 when their characteristics equal x. This is a special case of the global effect (3) when $\theta(x, v) = G(w, x, v)$. We construct the quantile structural function (QSF) in the selected population as the left-inverse of the mapping $w \mapsto G^s(w, x)$. That is:

$$q^s_{x_0}(\tau, x) = \inf\{w \in \mathbb{R} : G^s_{x_0}(w, x) \geq \tau\}.$$

The QSF gives the quantiles of $g(x, \varepsilon)$. Unlike $G^s_{x_0}(w, x)$, $q^s_{x_0}(\tau, x)$ cannot be obtained by integration of the corresponding local effect, $q(\tau, x, v)$, because we cannot interchange quantiles and expectations. The τ-quantile treatment effect of changing X from x_0 to x_1 in the selected population is:

$$q^s_{x_0}(\tau, x_1) - q^s_{x_0}(\tau, x_0),$$

which measures how the τ-th percent quantile of W changes among individuals in the selected sample with observed characteristics x_0 when their observed characteristics are set equal to x_1. In Section 4 we use quantile treatment effects to evaluate the impact of education on wages.

We also employ global effects to generate counterfactual distributions constructed by integrating the DSF with respect to different distributions of the explanatory variables X and the control variable V. We then perform wage decompositions (e.g., DiNardo, Fortin and Lemieux, 1996, Chernozhukov, Fernández-Val and Melly, 2013, Firpo, Fortin and Lemieux, 2011, Arellano and Bonhomme, 2017, and FVV 2018) to examine the underlying forces of changes in the hourly wage distribution. We focus on the global effects for the selected population. The decompositions are based on the following expression for the observed distribution of W:

$$G_W^s(w) = \int_{\mathcal{X}\mathcal{Z}\mathcal{V}} F_{W|X,V}^s(w|x,v) dF_{X,Z,V}^s(x,z,v),$$

where we use that W is independent of Z conditional on X, V and $H > 0$. The superscript s is again used for the selected sample and to simplify notation we drop the explicit conditioning on $H > 0$. From Bayes' rule and the properties of conditional probability, we rewrite this equation as:

$$G_W^s(w) = \frac{\int_{\mathcal{Z}} \int_{\mathcal{X}\mathcal{V}(z)} G(w,x,v) \mathbf{1}\{h(x,z,v) > 0\} dF_{X,V|Z}(x,v|z) dF_Z(z)}{\int_{\mathcal{Z}} \int_{\mathcal{X}\mathcal{V}(z)} \mathbf{1}\{h(x,z,v) > 0\} dF_{X,V}(x,v|z) dF_Z(z)}.$$

We construct counterfactual distributions by combining the component distributions G and $F_{X,Z,V}$ and the selection rule h from different populations that can correspond to different time periods or demographic groups. Thus, let G^t, and F_{X_k,Z_k,V_k} denote the distributions in groups t, r, and k, and let $\mathcal{X}\mathcal{Z}\mathcal{V}_k$ denote the joint support $\mathcal{X}$, $\mathcal{Z}$ and $\mathcal{V}$ for group k. Then, the counterfactual distribution of W where the wage structure G is as in group t, the joint distribution of the characteristics and the control function is as in group k, and the selection rule is as in group r is defined as:

$$G_{W_{\langle t|k,r \rangle}}^s(w) = \frac{\int_{\mathcal{Z}_r} \int_{\mathcal{X}\mathcal{V}_r(z)} G^t(w,x,v) \mathbf{1}\}h^r(x,z,v) > 0\} dF_{X_k,V_k|Z_k}(x,v|z) dF_{Z_k}(z)}{\int_{\mathcal{Z}_r} \int_{\mathcal{X}\mathcal{V}_r(z)} \mathbf{1}\{h^r(x,z,v) > 0\} dF_{X_k,V_k|Z_k}(x,v|z) dF_{Z_k}(z)}.$$

From the discussion above (*i.e.* Section 3.2), we know that $G^t(w,x,v)$ is identified for all combinations of x and v for which $(x, v) \in \mathcal{XV}_t$. The integration above is over the observed combinations of x and v of the population r, conditional on the outcome for Z equal to z. Hence, a necessary condition for the identification of $G^s_{W_{\langle t|k,r\rangle}}(w)$ is $\mathcal{XV}_r(z) \subseteq \mathcal{XV}_t$. Note that $\mathcal{XV}_r(z)$ is a subset of the combination of x and v in the total population of r, *i.e.* the identification set of the population r, or $\mathcal{XV}_r(z) \subseteq \mathcal{XV}_r$. Hence, a sufficient condition of the above integral to be identified is that $\mathcal{XV}_r \subseteq \mathcal{XV}_t$. This implies that it is sufficient that the observed combinations of x and v in population r must be a subset of the population t. That is, individuals with a certain combination of x and v in the selected sample for population r are also in the selected sample for population t. Under this definition, the observed distribution in group t is $G^s_{W_{\langle t|t,t\rangle}}$. We can write:

$$G^s_{W_{\langle t|k,r\rangle}}(w) = \frac{\int_{\mathcal{Z}_k} \int_{\mathcal{XV}_k(z)} G^t(w,x,v)\mathbf{1}\{(x,v) \in \mathcal{XV}_r(z)\} dF_{X_k,V_k|Z_k}(x,v|z) dF_{Z_k}(z)}{\int_{\mathcal{Z}_k} \int_{\mathcal{XV}_k(z)} \mathbf{1}\{(x,v) \in \mathcal{XV}_r(z)\} dF_{X_k,V_k|Z_k}(x,v|z) dF_{Z_k}(z)}$$

and since $(x, v) \in \mathcal{XV}_r(z)$ if and only if $v > \mathbb{P}(H_r \leq 0|X_r = x, Z_r = z)$, this equation has sample analog:

$$\widehat{G}^s_{W_{\langle t|k,r\rangle}}(w) = \frac{\sum_i \widehat{G}^t(w, X_{ik}, V_{ik})\mathbf{1}\{V_{ik} > \widehat{\mathbb{P}}(H_r \leq 0|X_r = X_{ik}, Z_r = Z_{ik})\}}{\sum_i \mathbf{1}\{V_{ik} > \widehat{\mathbb{P}}(H^r \leq 0|X_r = X_{ik}, Z_r = Z_{ik})\}}.$$

We can decompose differences in the observed distribution between group 1 and 0 using counterfactual distributions:

$$G^s_{W_{\langle 1|1,1\rangle}} - G^s_{W_{\langle 0|0,0\rangle}} = \underbrace{[G^s_{W_{\langle 1|1,1\rangle}} - G^s_{W_{\langle 1|1,0\rangle}}]}_{(1)} + \underbrace{[G^s_{W_{\langle 1|1,0\rangle}} - G^s_{W_{\langle 1|0,0\rangle}}]}_{(2)} + \underbrace{[G^s_{W_{\langle 1|0,0\rangle}} - G^s_{W_{\langle 0|0,0\rangle}}]}_{(3)}, \quad (6)$$

where (1) is a selection effect due to the change in the selection rule given the distribution of the explanatory variables and the control function, (2) is a composition effect due to the change in the distribution of the explanatory variables and the control function, and (3) is a structure effect due to the change in the conditional distribution of the outcome given the explanatory variables and control function. Finally, even though we do our decompositions at different points of the distribution, it is also possible to investigate decompositions for other objects. An example is the mean, which was the focus in earlier papers.

Prior to presenting our empirical results we precisely define these respective effects since they differ from definitions in the earlier literature due to the non-parametric nature of our model. First, the selection effect captures the difference between the actual distribution of wages and the counterfactual distribution if the individuals are selected as in year t, but the degree of selection is that of the base year. In the Heckman selection model, which focuses on the mean, this corresponds to the change in the value of the Mills ratio resulting either from changes in the coefficients from the first step or from changes in the explanatory variables appearing in that step. Second, our structural effect captures both the changes in the conditional distribution given the values of the explanatory variables and the changes in that distribution given the value of the control variable. The latter would be captured by the changes in the coefficient of the Mills ratio in the Heckman selection model and is attributed to the selection effect. This is reasonable in that context since this coefficient captures the correlation between the unobserved characteristics in the selection and outcome equations and reflects the selectivity of the sample. However, such an interpretation is not possible in the nonseparable model we consider here.

To illustrate this suppose (1) and (2) take the following parametric form:

$$W_t = \alpha_t + \beta_t \varepsilon_t \quad \text{if } H_t > 0$$
$$H_t = \max\{\gamma_t + \delta_t \eta_t, 0\},$$

where ε_t and η_t are dependent normal (Gaussian) random variables with zero mean and unit variance. When $\beta_t = \delta_t = 1$, this specification corresponds to the standard Heckman selection model (with unit variances). It can be shown that the mean of W_t in the selected population equals:

$$\mathbb{E}(W_t | H_t > 0) = \alpha_t + \beta_t \rho_t \lambda \left(\frac{\gamma_t}{\delta_t} \right),$$

where $\lambda(\cdot)$ and ρ_t respectively denote the Mills ratio and the correlation coefficient between ε_t and η_t. The earlier literature refers to the second component of this equation as the selection effect as it measures the difference between the average wage that is observed and the average wage that would have been observed if everybody would have been working. From the definitions in (6), the selection effect at the mean is:

$$\beta_t \rho_t \left[\lambda \left(\frac{\gamma_t}{\delta_t} \right) - \lambda \left(\frac{\gamma_0}{\delta_0} \right) \right], \tag{7}$$

while the structural effect at the mean is:

$$\alpha_t - \alpha_0 + (\beta_t \rho_t - \beta_0 \rho_0) \lambda \left(\frac{\gamma_0}{\delta_0} \right).$$

When $\beta_t = \delta_t = 1$, the sum of the selection effect (equation (7)) and the final term of this structural effect equals the change in the selection effect defined by Mulligan and Rubinstein (2008).[5], *i.e.*

$$\lambda(\gamma_t) - \lambda(\gamma_0) + (\rho_t - \rho_0) \lambda(\gamma_0).$$

Thus, a component of their selection effect is captured by the structural effect. When $\beta_t = \gamma_t = 1$, it is reasonable to describe this as a selection effect as it measures the changes in the type of selection. However, when $\beta_t \neq 1$, the second component also captures changes in the returns to the unobserved characteristics of individuals who worked in the base year.

4 Empirical Results

4.1 Hours Equation

The left panel of Figure 3 plots the employment rates of males and females, defined as the percentage of sample participants who report positive annual hours of work and positive annual earnings in the year before the survey. The male employment rate fluctuates cyclically around a downward trend starting at 90.0 percent in 1976, reaching a minimum of 82.1 percent in 2012, and increasing slightly to 83.1 percent in 2016. The female rate increases from a low of 56.5 percent in 1976 to a high of 75.3 percent in 2001. It is 70.0 percent in 2016.

The right panel of Figure 3 plots average annual hours worked for wage earners.

[5] As defined in equation (12) of their paper.

For males they vary cyclically around a slightly upward trend. They decrease from 2032 in 1976 to a sample minimum of 1966 hours in 1983. Then, they increase reaching a maximum of 2160 hours in 2001 before ending at 2103 hours in 2016. For females, average annual hours increase by nearly 20 percent between 1976 and 2000, from 1515 to 1792 hours. The increase continues after 2000, although more slowly, reaching 1838 hours in 2016. These patterns suggest important movements along both the intensive and the extensive margins of labor supply. The large variation in the average annual hours worked and the movements in the participation rates suggest that changes in the hourly wage distribution reported in Figure 1 may partially reflect compositional changes, in terms of both observable and unobservables, of the work force.

We first estimate the control function, defined as the conditional distribution function of hours, via a distribution regression of annual hours of work on conditioning variables for the whole sample. This first step includes those reporting zero hours of work. We do this via logistic regression, separately by gender for each year, as outlined in the appendix. The conditioning variables include age and age squared, dummy variables for the highest educational attainment reported (less than high school, high school, some college, or college or more), a dummy variable for marital status (married or not), a dummy variable for being nonwhite, a set of dummy variables for the region of residence (North-East, South, Central and West), the number of children in the household, the number of household members, a dummy variable for the presence of children aged less than 5 years and a dummy variable for the presence of unrelated individuals in the household. We also include the interaction of the educational dummies with age and age squared. Each of these variables appears in the hourly wage models except those capturing family composition. The family composition variables, along with the variation in hours, are the basis of identification.[6]

As our focus is on the estimation of the wage equation, we do not discuss the estimation results for the hours equation although we highlight the role of the exclusion restrictions. Given the nature of our selection rule the assumption that annual hours do not affect the hourly wage rate means that the variation in hours

[6] For a detailed discussion of issues related to identification in this model the reader is referred to FVV.

across individuals is a source of identification. That is, the variation in hours induces movement in the control function for the sample of workers. In addition we use measures of family household composition as predictors of hours which do not directly affect wages. While one can argue that household composition may affect hourly wage rates, we regard these restrictions as reasonable. We also highlight that similar restrictions have been employed in the earlier literature (see, for example, Mulligan and Rubinstein 2008).

4.2 Wage Equation and the Impact of Education

We now focus on the determinants of individual hourly wage rates. We estimate model (1) separately for males and females and each of the cross sections by distribution regression over the subsample of workers reporting a positive wage. The conditioning variables are those in the hours equation except for the household composition variables. We also include the value of the control function and its square, and the remaining conditioning variables are all interacted with the control function.

Substantial empirical evidence suggests that the increasing wage dispersion partially reflects changes in the impact of education (see for example, Autor, Katz and Kearney, 2008, and Murphy and Topel 2016). We explore this by estimating the impact of education on wages. We treat an individual's education level as exogenous and the "endogeneity/selection" to which we refer reflects the hours of work decision.

Education is represented by three dummy variables indicating that an individual has acquired, as her highest level of schooling, "high school", "some college" or "college or more". The excluded category is "less than high school". Since 1992, the CPS measures educational attainment by the highest year of school or degree completed rather than the previously employed "highest year of school attendance". Although the educational recode by the IPUM-CPS aims at maximizing comparability over time, there is a discontinuity between 1991 and 1992 in the way those with a high school degree and some college are classified. Figure 5 presents the fraction of wage earners with a high school degree or less and the fraction of those with at least a college degree. The figure illustrates the dramatic changes in educational attainment of the labor force. Male workers with at most a high-school degree fell steadily from 64.1 percent in 1976 to 38.4 percent in 2016, while those with at least a college degree rose from 20.4 percent in 1976 to 35.2 percent in 2016. The trends for females are even more striking as those with at most a high-school degree fell from 69.1 percent in 1976 to 29.7 percent in 2016, while those with at least a college degree rose from 16.1 percent in 1976 to 40.1 percent in 2016.

We estimate the various treatment effects introduced in Subsection 3.3 via the following sample analog of (4):

$$\frac{1}{n_e}\sum_{i=1,E_i=e}\widehat{\mu}(X_i,e',\widehat{V}_i)-\frac{1}{n_e}\sum_{i=1,E_i=e}\widehat{\mu}(X_i,e_i,\widehat{V}_i),$$

where E_i is education level of individual i, e and e' are the actual and counterfactual education levels respectively, X_i denotes other observed characteristics, and $\widehat{V}_i$ denotes the estimated control function for individual i. As discussed in the appendix, the function $\widehat{\mu}$ is the estimated LASF based on a flexibly specified regression of log wages on (X_i, E_i, V_i).[7] The quantile treatment effects can be estimated similarly although we employ the LSDF rather than the LASF to obtain global estimates of the structural distribution function. We invert that distribution to obtain the quantile treatment effects. As we use log wages, these effects can be interpreted as percentage changes. To satisfy the identification restriction we calculate the wage increase individuals with low levels of education, e, would receive if they achieved a higher level of education, e'. While this is the "opposite" of a treatment effect, we employ this terminology for simplicity.

Figures 6 to 8 present the average treatment effects for various contrasts in education levels. The average treatment effects for the "high school" to "less than high school" comparison, shown in Figure 6, reveal that for our sample period the effect is increasing but reasonably flat.

Figure 7 provides the estimated treatment effect for the contrast between "some college" and "less than high school". For males it is 33 percent in 1976 and 44.9 percent in 2016. It peaks at 50 percent in 2008. Although there are episodic declines, there is a sizable increase over the sample period. For females there is a notable increase during the 1980's although generally the pattern and magnitude of the

[7]FFV provide conditions under which the confidence intervals can be estimated via bootstraps.

changes are similar to those of males.

While Figures 6 and 7 are suggestive that education has contributed to inequality, the evidence in Figure 8 providing the treatment effect for the difference between "college" and "less than high school", is dramatic. Over the sample period this premium increases from 50 to 89 percent for males and from 69 to 88 percent for females. For males there is a steady increase with several large jumps. The occasional decreases appear to either reflect sampling issues or cyclical influences. The evidence for females is similar although there is a large decline in the late 1970's before large increases in both the 1980's and 1990's. The growth in the final decades of the previous millennium is not observed in the new millennium

While these various average treatment effects are informative about general trends, they fail to reflect the degree of heterogeneity in the impact of specific educational treatments. Accordingly, we provide corresponding quantile treatment effects in Figures 9 to 11. We present the quantile treatment effects at the 25th, 50th and 75th percentiles as these appear to be indicative of the lower, middle and upper parts of the wage distribution.[8]

Figure 9 presents the "high school"/"less than high school" comparison. Several features are notable. First, despite the growth in this premium for males and females from the period 1976 to 2000, the profiles are reasonably flat over large parts of the sample. Second, for males there are some differences across quantiles with those at the 75th percent quantile showing the strongest evidence of growth. Third, females have relatively similar profiles at the three quantiles we examine.

Figure 10 contrasts "some college" to "less than high school" and provides a clearer indication that this premium has increased over the sample with particularly strong evidence of large increases in the 1980's and part of the 1990's for females. The variability across the quantiles is small.

Figure 11 presents the quantile treatment effects for the college premium. There is a notable increase in the college effect for males at the first quartile until around 2000 before it flattens for the remainder of the sample period. At the median there is a more prolonged increase in the premium. This reflects a more heterogeneous effect with the quantile effect at the first quartile starting at around 52 percent

[8]The quantile treatment effects are based on location in the wage distribution. Thus the treatment effects are not necessarily increasing as the quantile at which they are evaluated increases.

before ending at 81 percent, while the median increases to 93 percent in 2013 before ending at 90 percent. The increase is even more evident at the third quartile.

Similar patterns appear for females although some additional features are interesting. There is dramatic growth at the median and below in the 1980's and parts of the 1990's and some leveling out over the 2000's. At the third quartile there is a steady increase over the 1980's and 1990's and periods of equally fast growth in some post 2000 periods. For females, there is significant heterogeneity in the effects across quantiles.

These estimated education treatment effects provide insight into the possible contributors to inequality. First, the returns to education have increased over our sample period and the premium for the higher education levels have increased remarkably. Second, the treatment effects, particularly for the college premium, show a greater degree of heterogeneity with the quantile treatment effects showing that the third quartile treatment effect for college education is very high for both males and females. Finally, the general trends and level of the average and quantile treatment effects are similar across gender although the trends in wages differ by gender. This suggests that the education premia is not the only contributing factor. While we do not attempt to isolate the role of the other conditioning variables they are included in the decomposition exercise. Note that these results are for the selectivity adjusted estimates. Although they are not reported here, the unadjusted estimates are similar. We also estimated local effects corresponding to different values of the control function. Those estimated effects did not reveal anything remarkable and are not reported here. The results are available from the authors.

4.3 Decompositions

For the decompositions we set the base years to 1976 for females and 2010 for males. The increasing female participation rate makes the choice of 1976 seem reasonable as it assumes that those individuals with a certain combination of x and v working in 1976 have a positive probability of working in any other year. By the same argument, a sensible choice of base year for males is 2010, the bottom of the financial crisis, as it has the lowest level of participation over our sample period. Nevertheless, this is somewhat harder to defend. The different base years for the genders means that we can compare trends but not wage levels across males and females. The decompositions are presented in Figures 12 to 16. They capture the impact of changes in the specified components on the change of the wage distribution. For example, the selection effect captures the change in the wage distribution due a change in the selection process.

We commence with the median as policy discussions frequently focus on this quantile. This is presented in Figure 14. During our sample period the male median wage decreases by 13 percent while that of females increased by 20 percent. These are similar to the total changes in the earlier figures. The total and structural effects are very similar for males with the small difference due to the composition effect. There is no evidence of any change in the impact of selection on the median wage. The composition effect is increasing and most likely reflects the increasing educational attainment of the workforce. The composition effect increases the median wage by around 2 percent. In contrast, the structural component, which appears to be strongly procylical, produces large negative effects. While there appear to be some upturns, coinciding with periods of improving economic conditions, the structural effect is negative over the entire period. This is a striking result. Our results are consistent with the paper of Chernozhoukov, et al. (2013) which performs a similar exercise for the period 1979 to 1988 without correcting for sample selection.

The median female wage increases by around 20 percent and Figure 14 reveals that this is entirely driven by the composition effect. The structural effect is negative for the whole period, but is less substantial than that for males. The turning points appear to coincide across genders. The contribution of the selection component is negative but negligible. This contrasts with Mulligan and Rubinstein (2008), who find a large change in the selection effect at the mean over time. Specifically, they find the selection effect changes from negative and large to positive and large. We do not find evidence for such a drastic effect, but as highlighted above, our selection effect differs from the conventional definition employed by Mulligan and Rubinstein (2008). That is, we incorporate the changes in the correlation between the unobservables in the outcome and selection equations into the structural effect and do not incorporate it in the selection effect. If the Mulligan and Rubinstein assertion that the correlation between the error terms increased over time is correct, implying that the selected sample of females became increasingly more productive relative to the total population, then our selection effect might underestimate the total change in the wage distribution due to changes in the selected sample over time.

We do not, however, find support for the conclusion of Mulligan and Rubinstein that the least productive females, in terms of unobservables, were working in the 1970's and 1980's. Our selection effect measures the difference between the observed distribution in any given year and the resulting distribution if the "least likely to participate" females among the selected sample would not have participated. If the "least likely females" to participate would have been the most productive, this would reduce the wage in the year of evaluation and reflect positive change in the selection effect in our figures for the 1970's and the 1980's. We do not find any support for a positive change in the selection effect.

The decompositions for the median suggests that the decline in the median male wage is due to the prices associated with the male skill characteristics. However, the evidence above established that the returns to schooling have generally increased and the male labor force has become increasingly more educated. This suggests that the returns to those individuals with the lowest levels of education must have markedly decreased. A similar pattern is observed for the female median wage although the less substantial negative structural impact is offset by the composition effects.

Figures 12 and 13 report the decompositions at the 10th and 25th percentiles and they are remarkably similar for males. They suggest the reductions in the male wage rate capture the prices associated with the human capital of individuals located at these lower quantiles. There is evidence of a negative structural component of around 25 to 30 percent at each of these quantiles in both the late 1990's and the late 2010's. While these effects are somewhat offset by the composition effects, the overall effect on wages is negative. The wage reduction appears to be driven by how worker's characteristics in this part of the wage distribution are valued. There are no signs of selection effects for males. Note that our results at the first decile for the period 1979 to 1988 appear to differ to those in Chernozhoukov et al. (2013). However, that study breaks the structural effect into separate components due to changes in the mandatory minimum wage, unionization, and the returns to any other characteristic. Not surprisingly, changes in the mandatory minimum wage have a large impact on the 10th percent quantile. Our estimates of the structural effect combines these various components and also includes the variation in the prices of unobservables. Our results are consistent with their study at any other quantile.

The evidence at the 10th and 25th percentiles for females is very different to that for males. First, there are greater differences between the 10th and 25th percentiles. The negative structural effects for females are more evident at the 10th percentile. The negative structural effects are small at the 25th percent quantile and offset by the composition effects. For both the 10th and 25th percent quantiles the overall wage changes become positive in the late 1990's and generally increase over the remainder of the sample.

Figures 15 and 16 present the decomposition of the wage growth at the 75th and 90th percentiles. The male wage at the third quartile shows a small increase. The structural component displays a similar pattern to that for females at the lower quantiles discussed above. That is, initially there is a large decrease before rebounding and remaining relatively flat from the early 2000's. Unlike the lower quantiles the negative changes resulting from the structural component are not sufficiently large to dominate the positive composition effects so the overall wage growth at the 75th percent quantile is positive from the early 2000's onwards.

The changes in the female hourly wage rate at the 75th percent quantile highlights our introductory discussion that the larger movements have occurred at the higher quantiles of the wage distribution. The changes in the structural component are initially negative before turning positive around the mid 1980's. From the beginning of the 1980's the positively trending structural component combines with the composition effect to produce a steadily increasing wage. However, while the decomposition at the 75th quantile suggests that the structural components are an important contribution to inequality at higher quantiles of the distribution, the results at the 90th percent quantile are even more supportive of this perspective. For males the structural component is less negative than at lower quantiles and this combined with the positive composition effect produces a wage gain for the whole period. For females the structural component is positive from the middle of the 1980's and has a larger positive effect than the composition effect. The two effects combine to produce a remarkable 41 percent growth in the wage.

Consider now the selection effects and recall that they capture the changes in the selection rule over time while assuming that the same unobservables have the impact on the hours of work decision as the year being evaluated. Thus, there can only be a change in the selection effect if the selection rule, and hence the employment rate, has changed over time. Given this interpretation, consider the role of selection for males. Figures 14 to 16 suggest that selection effects cannot explain the observed changes in the males' wage distribution. At the 50th, 75th and the 90th percentiles the selection effect is essentially zero. This result is not surprising as males in this area of the wage distribution have a strong commitment to the labor force. Moreover, not only is there likely to be relatively little movement on the extensive margin, there is also likely little movement on the intensive margin given males' level of commitment to full-time employment. One might suspect that it would be more likely to uncover changes in the selection effects at the lower parts of the wage distribution as these individuals are likely to have a weaker commitment to full-time employment and thus the movements at the extensive and intensive margins due to unobservables may be more important. However, the evidence does not support this.

Now focus on the selection effects for females. At higher quantiles there is little evidence of changes in selection. However, similar to males, females located in this part of the wage distribution are likely to have had a relatively strong commitment to employment in 1976 and thus there were no substantial moves in their hours distribution. However, at the 10th and 25th percentiles the selection effects seem economically important.[9] For example, at the 10th percentile in 2016 the selection effect contribution is 2.2 percent, while the total wage change is between 8 and 9 percent. Thus the female wage was lowered by 2 percent due to the increased participation of females in the later years of our sample period. This is consistent with the "positive selection" finding of Mulligan and Rubinstein (2008) for the 1990's. We also find a similar relationship for the late 1970's and 1980's. Our results generally suggest a positive relationship between the control variable V and wages at the bottom of the distribution. This implies those with the highest number of working

[9]The confidence intervals for these selection effects are presented in Figure 17. They indicate that for many time periods they are statistically significantly different from zero.

hours, after conditioning on their observed characteristics, had the highest wages. The trends implied by our results suggest that selection becomes more important as we move further down the female wage distribution. This is similar to the findings of Arellano and Bonhomme (2017) and the empirical evidence in FVV which both study the evolution of female wages in the British labor market for the period 1978 to 2000.

5 Discussion

A number of our empirical results are notable. The impact of the educational treatment varies drastically by level of attainment. While the return to completing high school and obtaining some college, relative to not completing high school, clearly increase wages, they do not appear to be major driving forces of increasing inequality. In contrast the college premium for both genders has increased dramatically over the sample period and has important implications for inequality. This is consistent with several other studies dating back to Murphy and Welch (1992).

The decompositions reveal a number of findings. The mechanisms driving wages differ by location in the wage distribution and across gender. For males, the fall in wages at the median and below appears partially due to the penalty associated with lack of education and other forces which are negatively affecting the lower skilled. This is reflected by the large negative structural effects for this area of the wage distribution. As more workers are receiving higher education the composition effects are positive and somewhat offset the negative structural effects. However, the large, and increasing educational premium, signals that the “penalty” to not being educated has increased. The negative structural effects for males are not restricted to the lower part of the wage distribution. For females, there are wage increases at each quantile we examined and these reflect, in part, positive and increasing composition effects. Moreover, while the structural effects are generally negative for the whole period at the 10th and 25th percentiles, they are typically positive at the quantiles we examined above the median. Most notably, the structural effects for females are not dampening wage growth to the same extent as for males and at some quantiles these structural effects are even substantial contributors to wage growth. At the lower parts of the female wage distribution the impact of selection is negative and can be substantial. Selection effects become more important as we move down the wage distribution. The patterns at the 10th and 25th percentiles are suggestive of even larger selection effects at lower percentiles.

The nature of our investigation does not provide direct insight into the macro factors generating these wages profiles. Nor does it provide evidence on the role of institutional factors which disproportionately influence certain sectors of the work force. However it does seem that the mechanisms affecting the wages at the bottom are very different than those influencing the top. At the top the evidence is supportive of an increasing skill premium. At the bottom it appears the prevailing considerations are those associated with the lack of protection of lower wage workers. These include the decreases in the real value of the minimum wage, the reduction in unionization and the union premium, and increases in employer bargaining power.

We now directly examine the issue of inequality. The 90/10th percentile ratios for hourly wages for males and females have increased by 55 and 38 percent respectively. Our evidence illustrates that hourly wage growth at different locations in the wage distribution is affected differently by the relevant factors. However, we are unable to directly infer from that evidence the respective contribution of these factors to the changes in inequality. For males recall that wage changes at the 10th percentile were due to the large negative structural effects. There was no evidence of selection and a small positive composition effect offset the decreases from around the early 1990's. The large gains at the 90th percent quantile reflected a steadily increasing composition effect and a structural effect which contributed both negatively and positively over the sample period. There are no signs of changes in selection effects. Figure 18 provides a decomposition of the changes in the 90/10 ratio for males. Unsurprisingly, the large increase in the 90/10 ratio for males is almost entirely due to structural effects, caused by the negative structural effects at the 1st decile and not positive effects at the ninth decile. The composition effects increase inequality via their large positive contribution at the ninth decile.

For females the evidence is more difficult to interpret. There is an increasingly positive composition effect at the first decile but the negative structural effect produces a decline in wages. At the 9th decile, there is a steadily increasing composition effect and a structural effect which is generally increasing wages. The large increase in the total effect for the majority of the sample period reflects the sum of these two positive effects. At the first decile, the evidence suggests that the changes in the selection effects are negatively affecting wage growth while there is no sign of selection at the 9th decile. Figure 18 presents the decomposition of the change in the 90/10 ratio for females. Given the various issues related to wage growth, it is not surprising that the change in the 90/10 ratio appears to reflect almost entirely a structural effect. The composition effect has slightly decreased this measure of inequality due to the large composition effects at the lower part of the wage distribution. The change in the selection effects has increased the 90/10 ratio. Moreover, its contribution in some periods is a relatively large fraction of the total change.

Although our primary focus is not gender inequality, we examine the trends in the male/female hourly wage ratio at different points of the wage distribution. These are reported in Figure 19. First, at all locations of the wage distribution females appear to be catching up. Moreover, the greatest gains appear at the median and below. This result should be treated with caution as it appears largely due to the reduction in the male wage and not large increases in the female wage. This is confirmed by a re-examination of Figure 1. Second, the improvement in the relative performance of females is almost entirely due to structural effects at all of the quantiles reported in Figure 19. As the earlier evidence suggested the sign and the size of the structural effects on the individual gender specific wages varied by sample period and location in the wage distribution, it is surprising that the impact on the gender wage ratios is so clear. The evidence suggests a relative improvement in the value of female labor at all points of the wage distribution. Third, the composition effects also have steadily increased the relative performance of females at all quantiles noting that the size of the effect diminishes as we move up the wage distribution. The effect is small at the 9th decile. Finally, the selection effects are increasing gender inequality although they only appear at the lower parts of the wage distribution.

6 Conclusions

This paper documents the changes in female and male wages over the period 1976 to 2016. We decompose these changes into structural, composition and selection components by implementing an estimation procedure for nonseparable models with selection. We find that male real wages at the median and below have decreased over our sample period despite an increasing skill premium and an increase in educational attainment. The reduction is primarily due to large decreases of the wages of the individuals with a low level of education. Wages at the upper quantiles of the distribution have increased drastically due to a large and increasing skill premium and this has combined with the decreases at the lower quantiles to substantially increase wage inequality. Female wage growth at lower quantiles is modest although the median wage has grown steadily. The increases at the upper quantiles for females are substantial and reflect increasing skill premia. These changes have resulted in a substantial increase in female wage inequality. As our sample period is associated with large changes in the participation rates and the hours of work of females we explore the role of changes in "selection" in wage movements. We find that the impact of these changes in selection is to decrease the wage growth of those at the lower quantiles with very little evidence of selection effects at other locations in the female wage distribution. The selection effects appear to increase wage inequality.

References

[1] Acemoglu D., and Autor D. (2011), "Skills, tasks and technology: Implications for employment and earnings" In D. Card and O. Ashenfelter (eds.), *Handbook of Labor Economics*, Vol. 4B: 1043–1171. Amsterdam: Elsevier Science, North-Holland.

[2] Amemiya T. (1984), "Tobit models: A survey", *Journal of Econometrics*, **24**, 3–61.

[3] Angrist J., Chernozhukov V., and Fernàndez-Val I. (2006), "Quantile regression under misspecification, with an application to the U.S. wage structure", *Econometrica*, **74**, 539–63.

[4] Arellano M., and Bonhomme S. (2017), "Quantile selection models with an application to understanding changes in wage inequality", *Econometrica*, **85**, 1–28.

[5] Autor D.H., Katz L.F., and Kearney M.S. (2008), "Trends in U.S. wage inequality: Revising the revisionists", *Review of Economics and Statistics*, **90**, 300–23.

[6] Bollinger C.R., Hirsch B.T., Hokayem C. M., and Ziliak J.P. (2018), "Trouble in the tails? What we know about earnings nonresponse thirty years after Lillard, Smith, and Welch", *Journal of Political Economy*, forthcoming.

[7] Borjas G. (1980), "The relationship between wages and weekly hours of work: The role of division bias", *Journal of Human Resources*, **15**, 409–23.

[8] Chen S., (1997) "Semiparametric estimation of the Type-3 Tobit model", *Journal of Econometrics*, **80**, 1–34.

[9] Chernozhukov V., Fernández-Val I., and Melly B. (2013), "Inference on counterfactual distributions", *Econometrica*, **81**, 2205–68.

[10] Danziger S., and Gottschalk P. (1995), *America Unequal*. Cambridge, MA: Harvard University Press.

[11] DiNardo J., Fortin N.M., and Lemieux T. (1996), "Labor market institutions and the distribution of wages , 1973–1992: A semiparametric approach", *Econometrica*, **64**, 1001–44.

[12] Fernández-Val I., van Vuuren A., and F. Vella (2018), "Nonseparable sample selection models with censored selection rules", Working paper, Boston University.

[13] Flood S., King M., Ruggles S., and Warren J. R. (2015), "Integrated public use microdata series, Current Population Survey: Version 4.0 [Machine-readable database]", University of Minnesota, mimeo.

[14] Foresi S., and Peracchi F. (1995), "The conditional distribution of excess returns: An empirical analysis", *Journal of the American Statistical Association*, **90**, 451–66.

[15] Fortin N.M., Lemieux T., and Firpo S. (2011), "Decomposition methods in economics." In D. Card and O. Ashenfelter (eds.), *Handbook of Labor Economics*, Vol. 4A: 1–102. Amsterdam: Elsevier Science, North-Holland.

[16] Heckman J.J. (1974), "Shadow prices, market wages and labor supply", *Econometrica*, **42**, 679–94.

[17] Heckman J.J. (1979), "Sample selection bias as a specification error", *Econometrica*, **47**, 153–61.

[18] Honoré B., Kyriazidou E., and Udry C. (1997), "Estimation of type 3 tobit models using symmetric trimming and pairwise comparison", *Journal of Econometrics*, **76**, 107–28.

[19] Imbens G. W., and Newey W.K. (2009), "Identification and estimation of triangular simultaneous equations models without additivity", *Econometrica*, **77**, 1481–1512.

[20] Juhn C., Murphy K.M., and Pierce B. (1993), "Wage inequality and the rise in returns to skill", *Journal of Political Economy*, **101**, 410–42.

[21] KATZ L. F., AND AUTOR D. H. (1999), "Changes in the wage structure and earnings inequality." In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. 3A: 1463–1555. Amsterdam: Elsevier Science, North-Holland.

[22] KATZ L.F., AND MURPHY K.M. (1992), "Changes in relative wages, 1963–87: Supply and demand factors", *Quarterly Journal of Economics*, **107**, 35–78.

[23] KOENKER R., AND BASSETT G. (1978), "Regression quantiles", *Econometrica*, **46**, 33–50.

[24] KRUEGER A., AND POSNER E. (2018), "A proposal for protecting low-income workers from monopsony and corruption." In J. Shambaugh and R. Nunn (eds.), *Revitalizing Wage Growth. Policies to Get American Workers a Raise*, 139–156. Washington, DC: Brookings.

[25] LARRIMORE J., BURKHAUSER R.V., FENG S., AND ZAYATZ L. (2008), "Consistent cell means for topcoded incomes in the Public Use March CPS (1976–2007)", *Journal of Economic and Social Measurement*, **33**, 89–128.

[26] LEE D.S. (1999), "Wage inequality in the United States during the 1980s: Rising dispersion or falling minimum wage?", *Quarterly Journal of Economics*, **114**, 977–1023.

[27] LEE M.J., AND VELLA F. (2006), "A semiparametric estimator for censored selection models with endogeneity", *Journal of Econometrics*, **130**, 235–52.

[28] LEMIEUX T. (2006), "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?" *American Economic Review*, **96**, 461–98.

[29] LILLARD L., SMITH J.P., AND WELCH F. (1986). "What do we really know about wages? The Importance of nonreporting and Census imputation", *Journal of Political Economy*, **94**, 489–506.

[30] MEYER B.D., MOK W.K.C., AND SULLIVAN J.X. (2015), "Household surveys in crisis", *Journal of Economic Perspectives*, **29**, 199–226.

[31] Mulligan C., and Rubinstein Y. (2008), “Selection, investment, and women’s relative wages over time”, *Quarterly Journal of Economics*, **123**, 1061–1110.

[32] Murphy K.M., and Topel R.H. (2016), “Human capital investment, inequality, and economic growth”, *Journal of Labor Economics*, **34**, S99–S127.

[33] Murphy K.M., and Welch F. (1992), “The structure of wages”, *Quarterly Journal of Economics*, **57**, 285–326.

[34] Newey, W.K. (2007), “Nonparametric continuous/discrete choice models” *International Economic Review*, **48**, 1429–39.

[35] Vella, F. (1993), “A simple estimator for models with censored endogenous regressors” *International Economic Review*, **34**, 441–57.

[36] Welch F. (2000), “Growth in women’s relative wages and inequality among men: One phenomenon or two?”, *American Economic Review Papers & Proceedings*, **90**, 444–49.

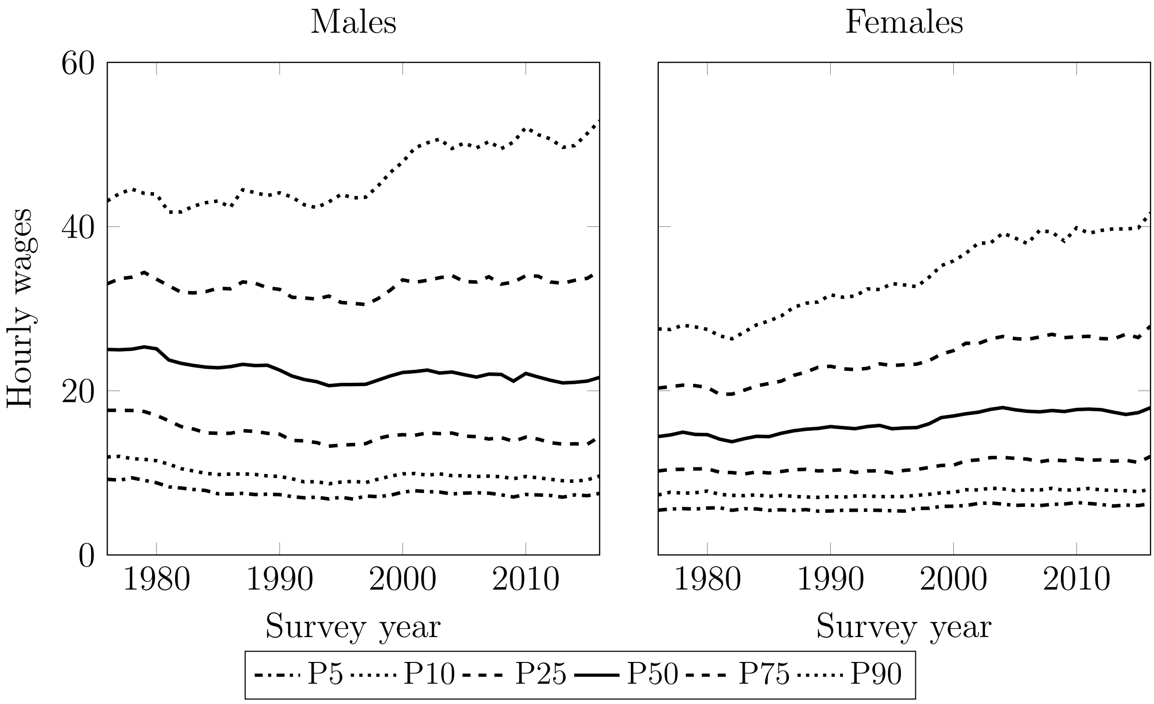

Figure 1: Percentiles of real hourly wages

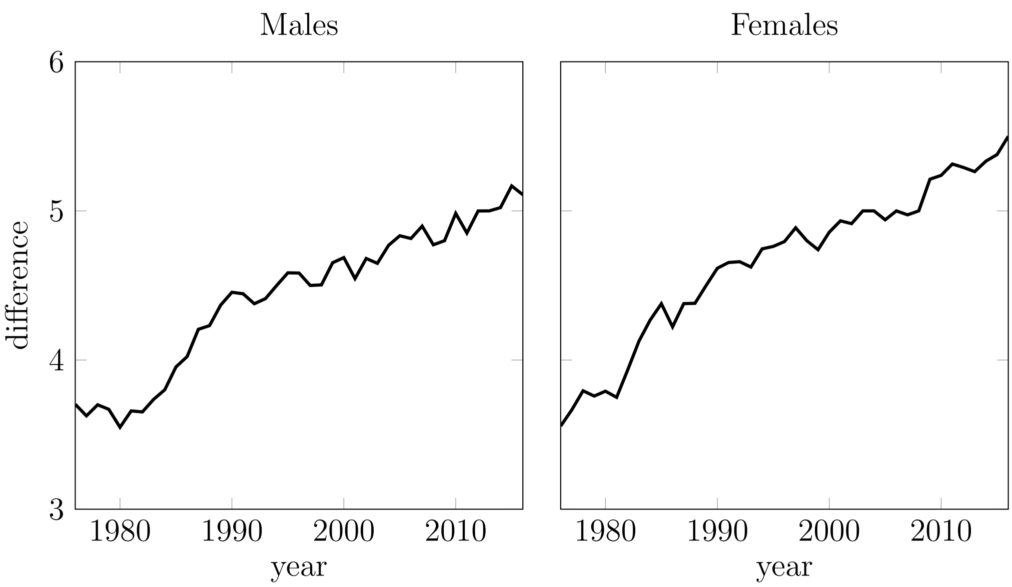

Figure 2: Development of the D9-D1 ratio.

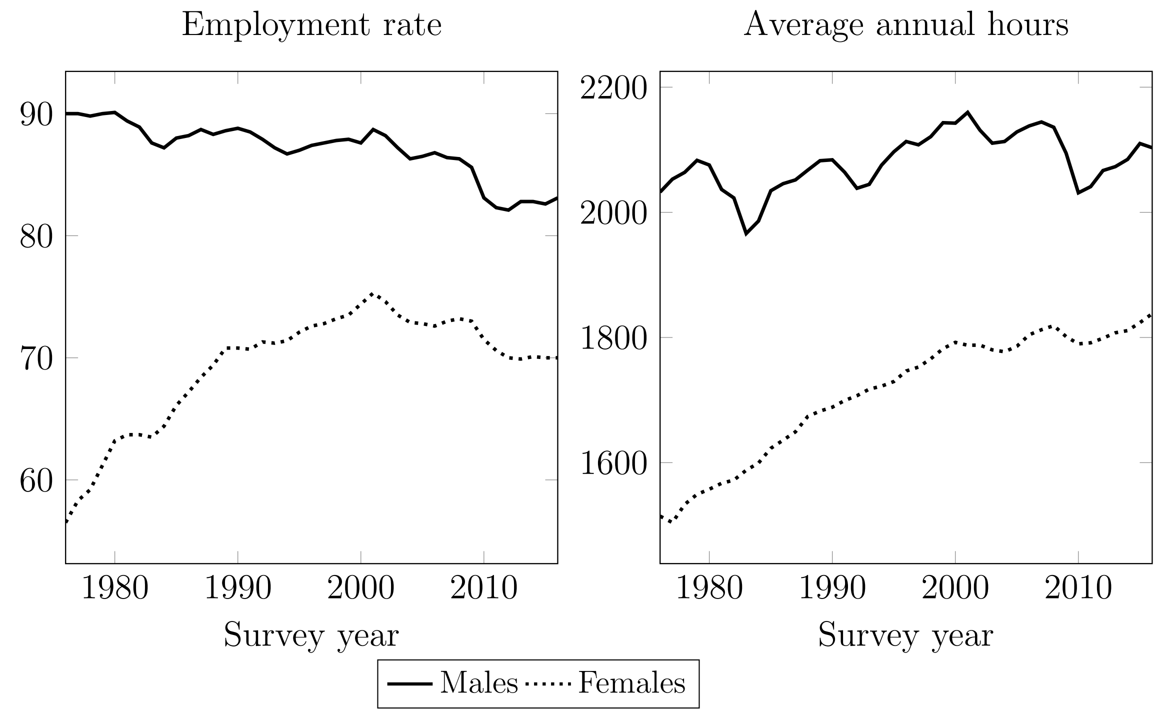

Figure 3: Employment rate and average annual hours worked of employees

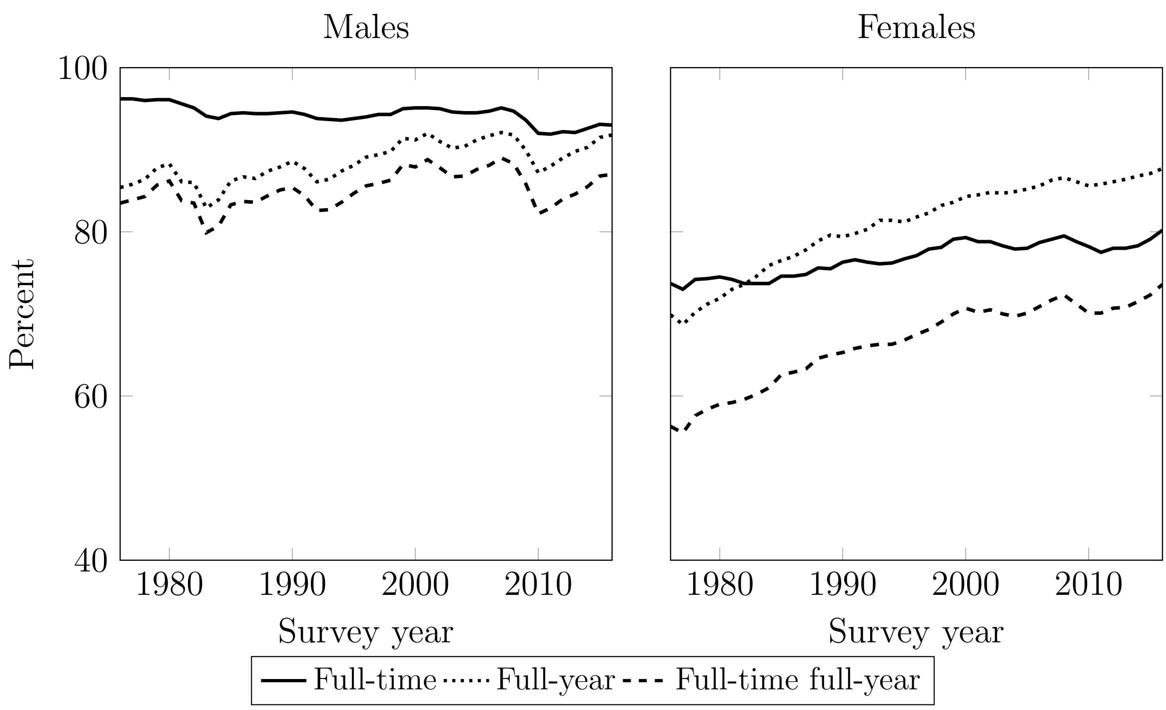

Figure 4: Fraction of wage earners working either full-time or full-year

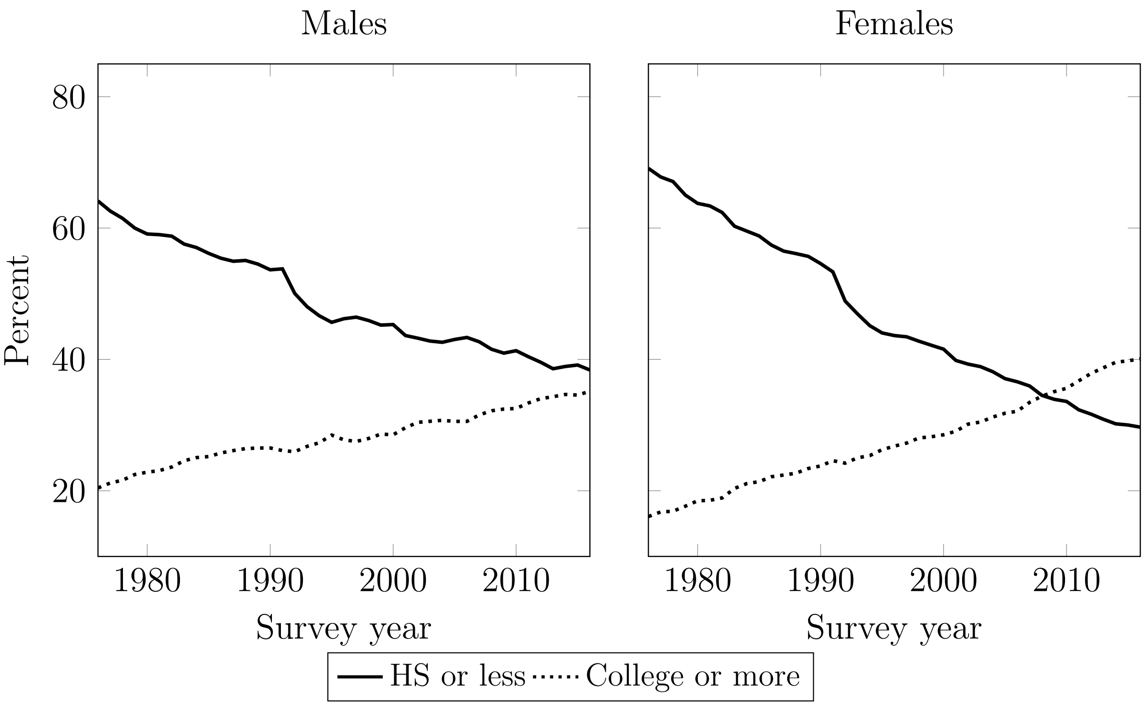

Figure 5: Distribution of wage earners by education level

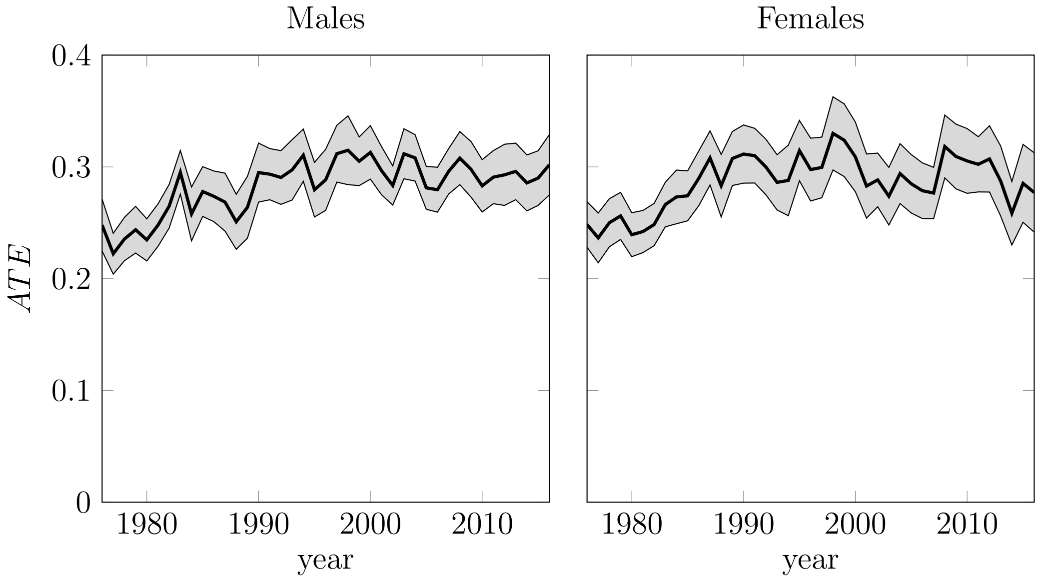

Figure 6: Average treatment effect of education, high school versus less than high school with correction.

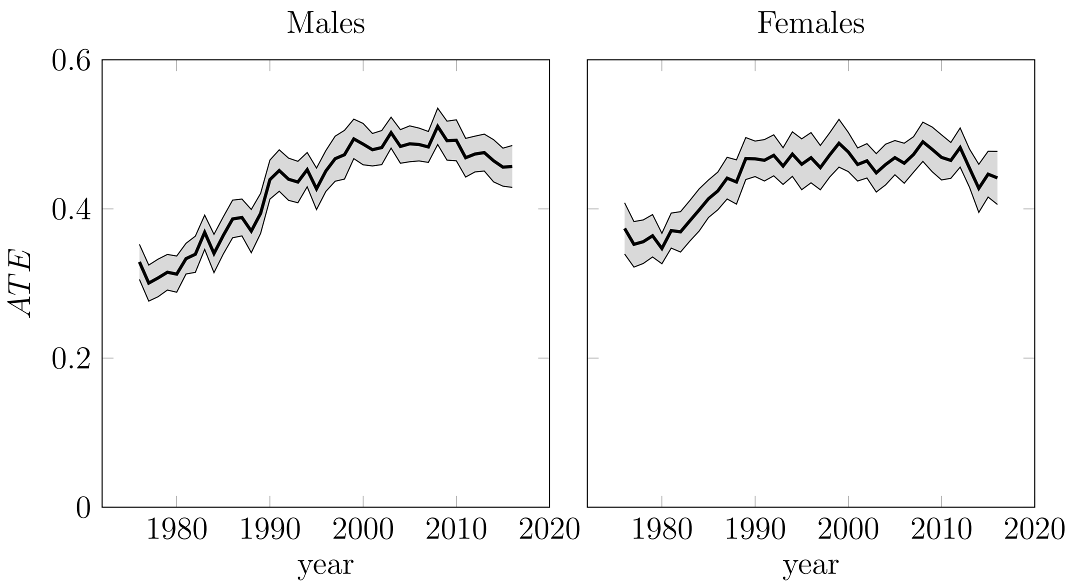

Figure 7: Average treatment effect of education, some college versus less than high school with correction.

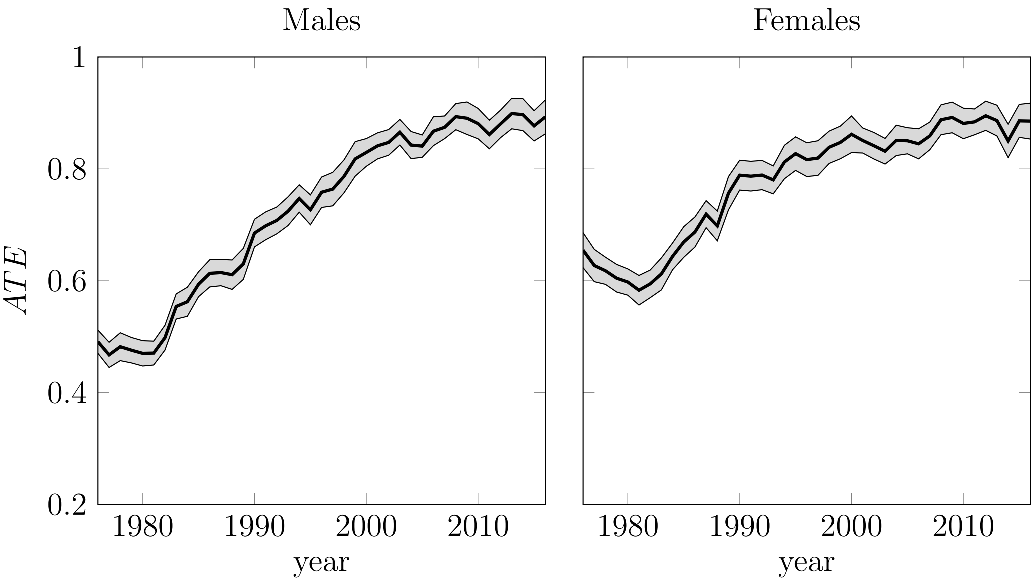

Figure 8: Average treatment effect of education, college versus less than high school with correction.

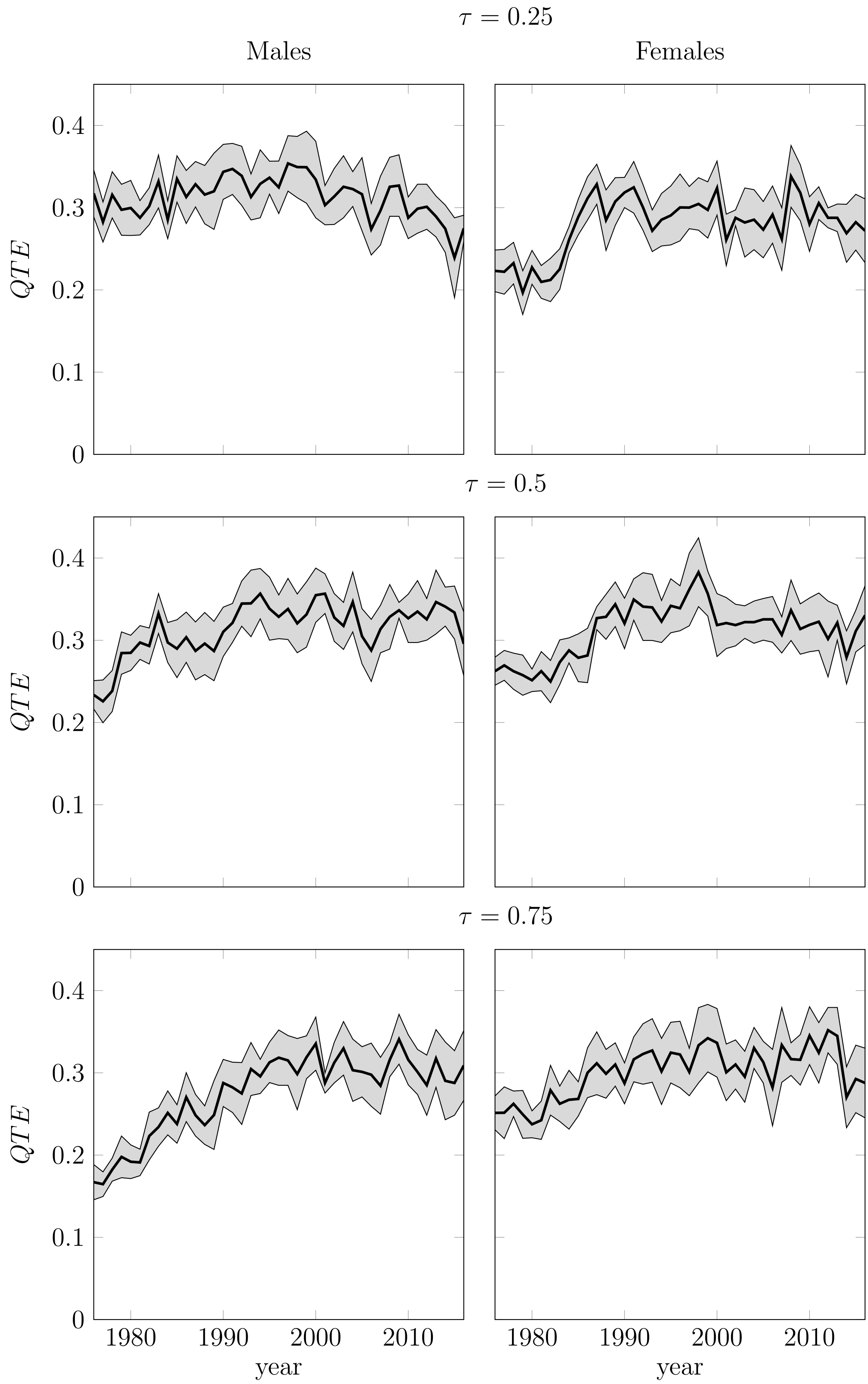

Figure 9: Quantile treatment effect of education, high school versus less than high school with correction.

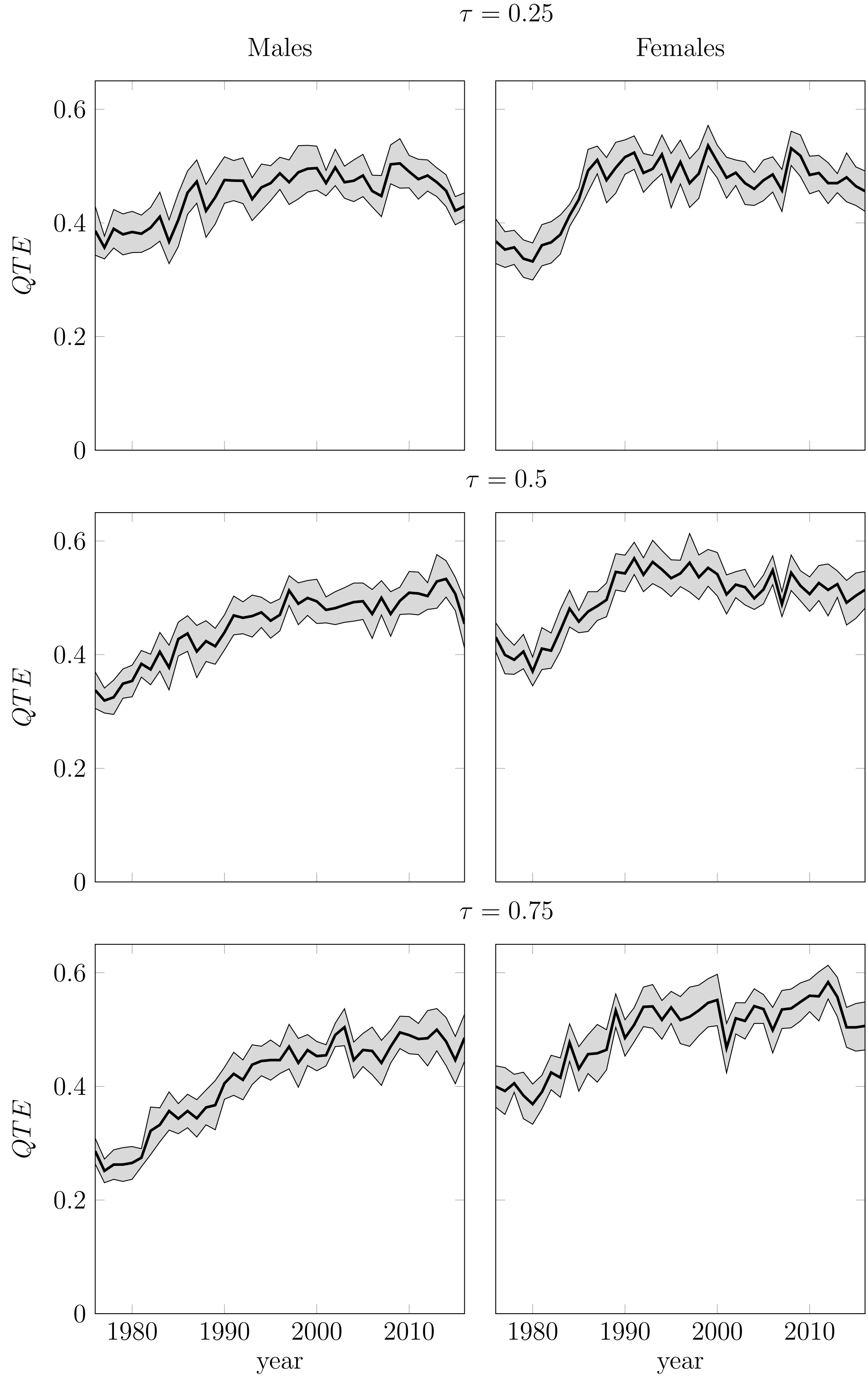

Figure 10: Quantile treatment effect of education, some college versus less than high school with correction.

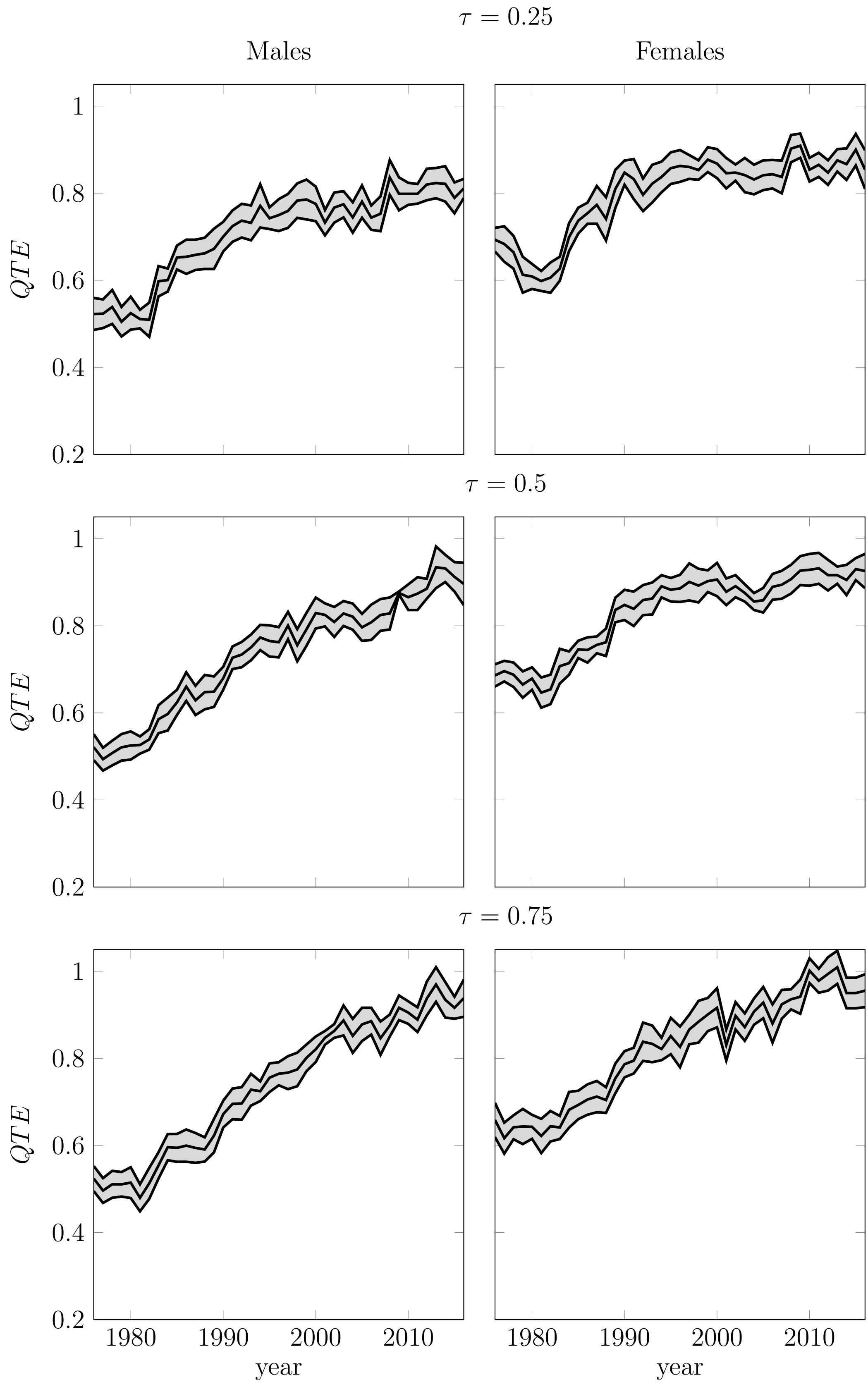

Figure 11: Quantile treatment effect of education, college versus less than high school with correction.

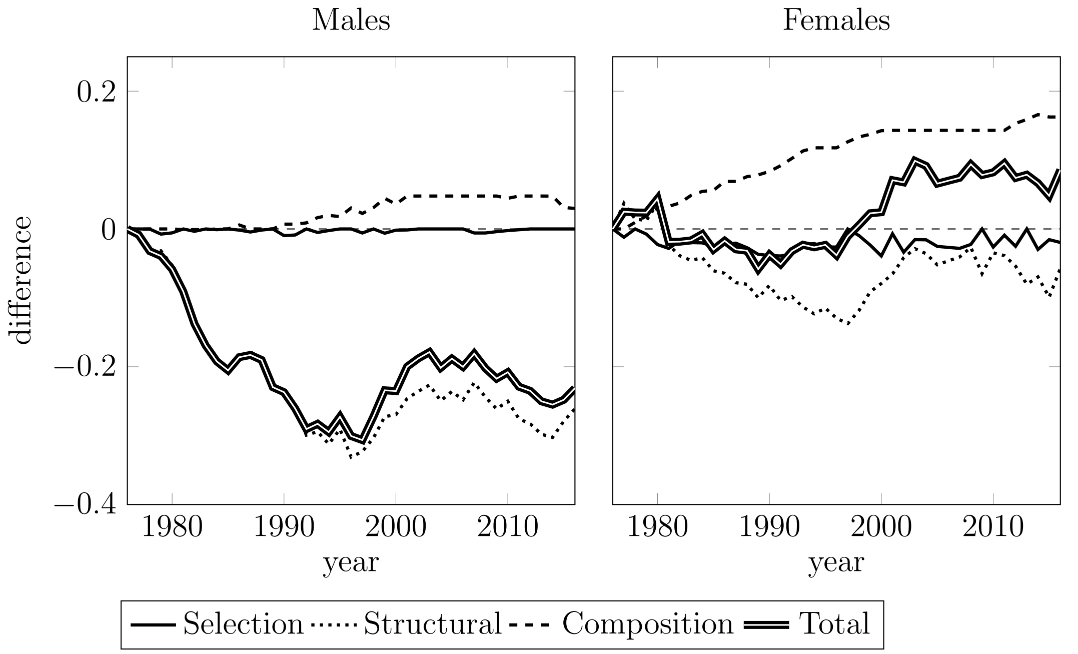

Figure 12: Decompositions at D1.

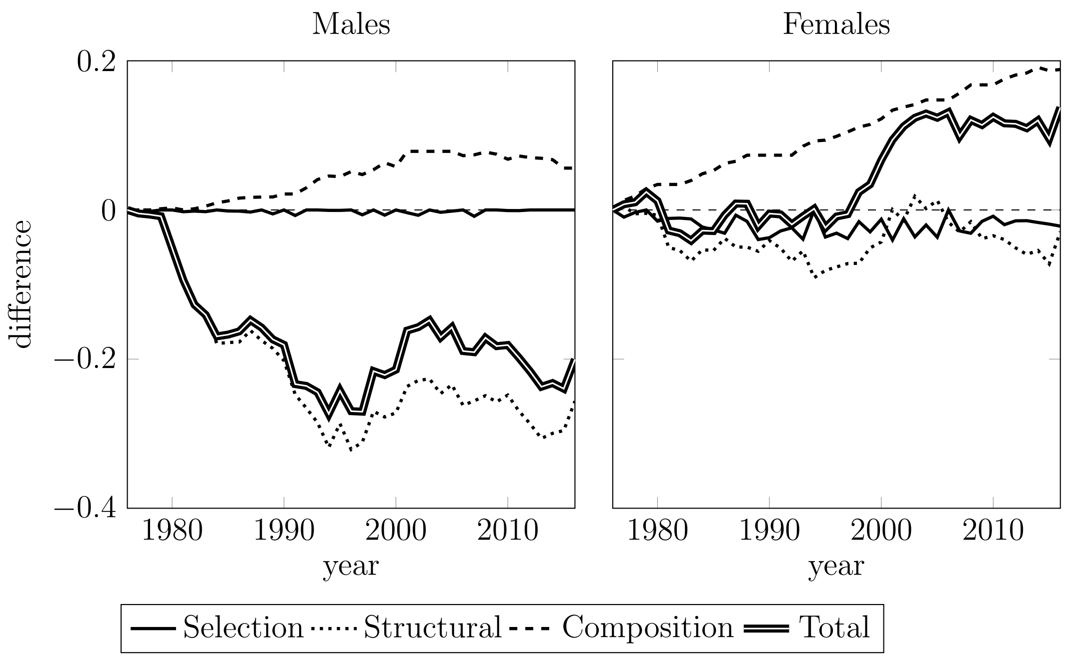

Figure 13: Decompositions at Q1.

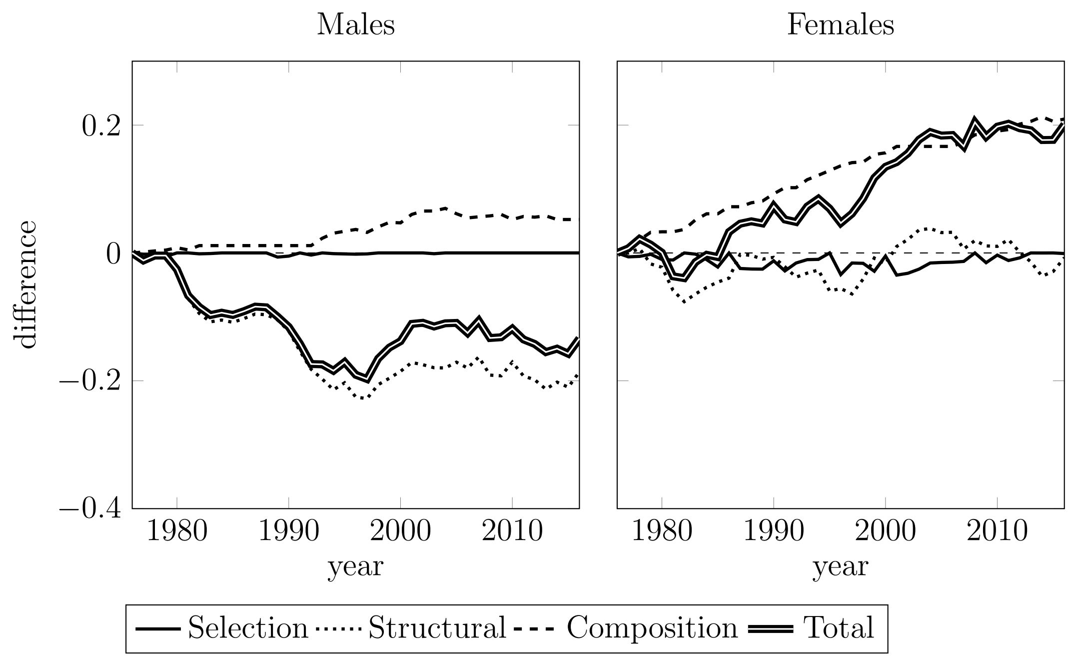

Figure 14: Decompositions at Q2.

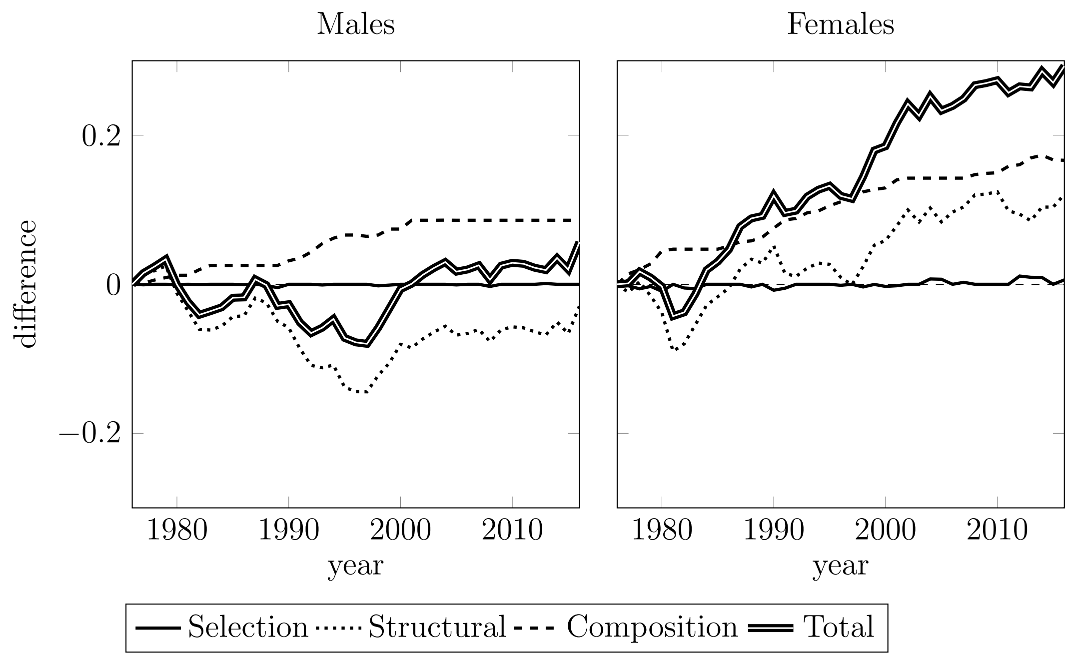

Figure 15: Decompositions at Q3.

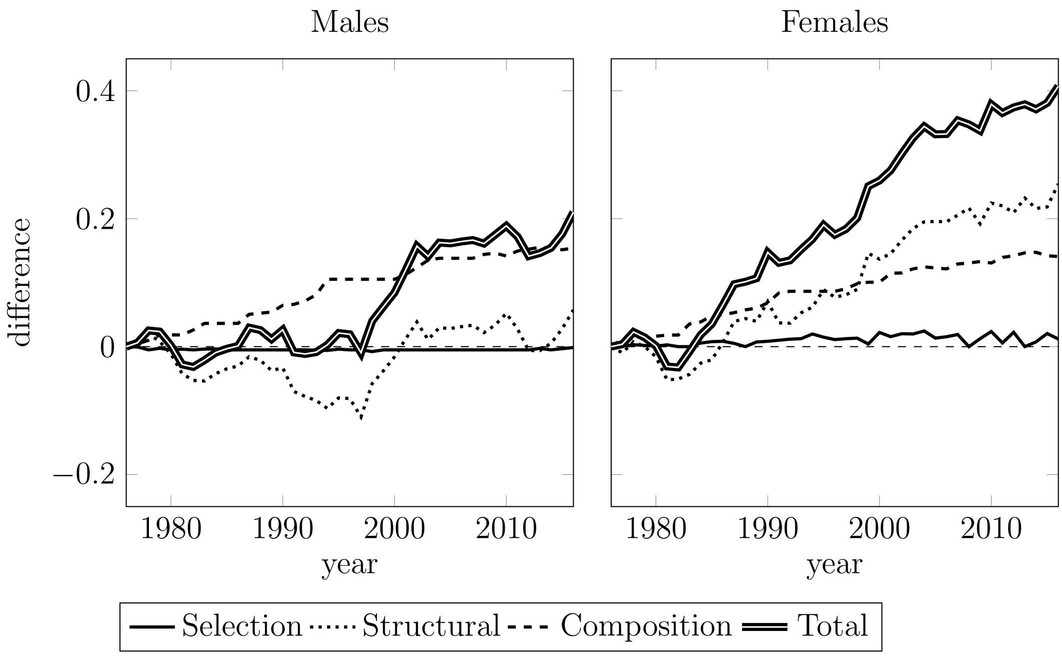

Figure 16: Decompositions at D9.

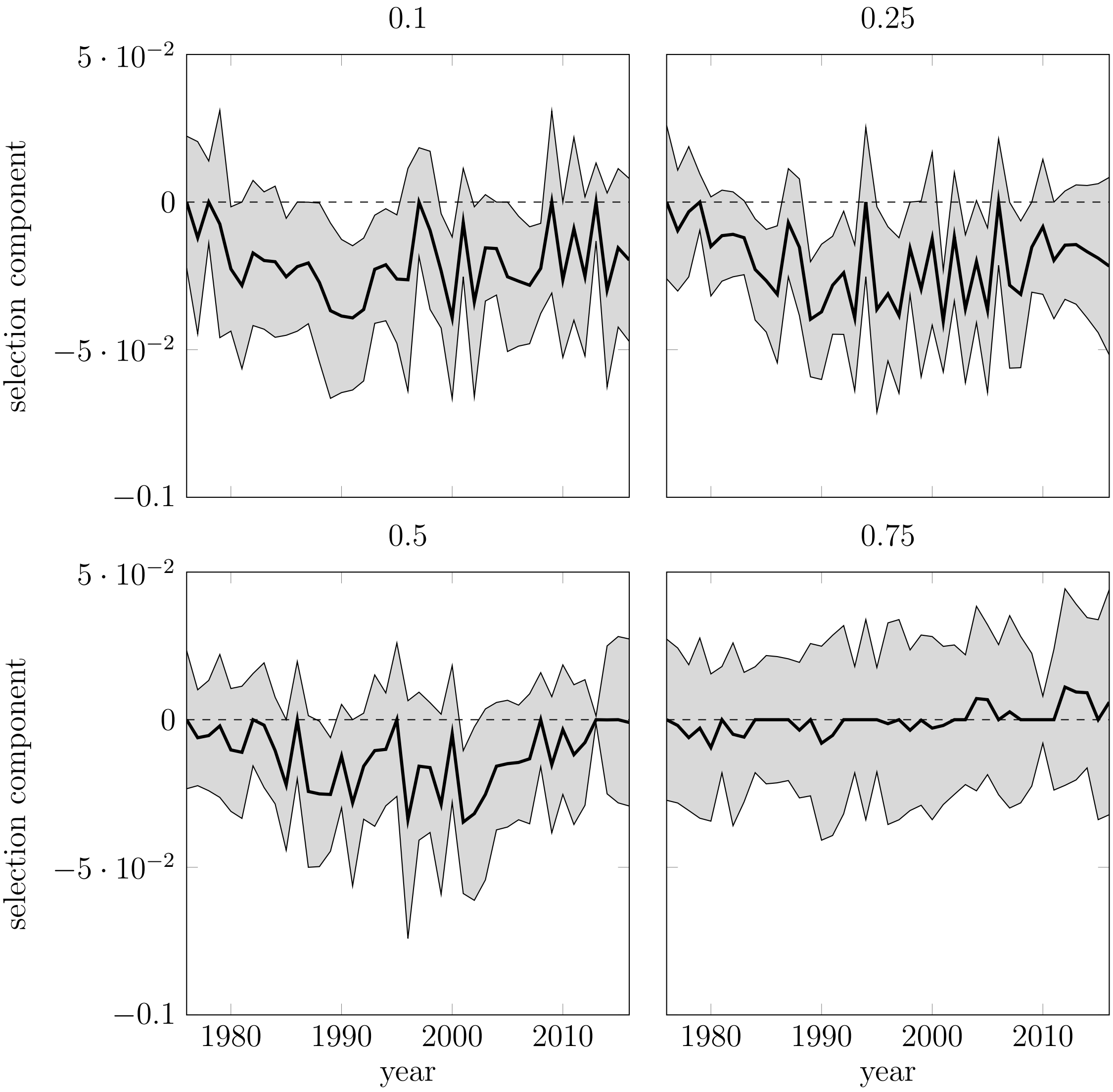

Figure 17: Selection component and 95% confidence intervals for females.

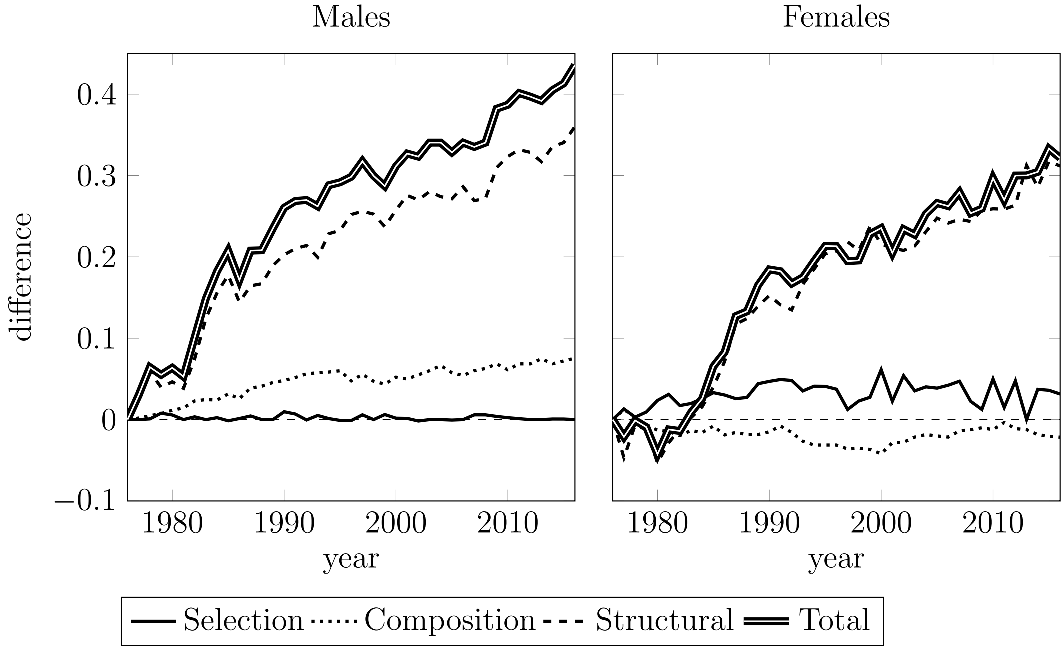

Figure 18: Decompositions at D9-D1 ratio.

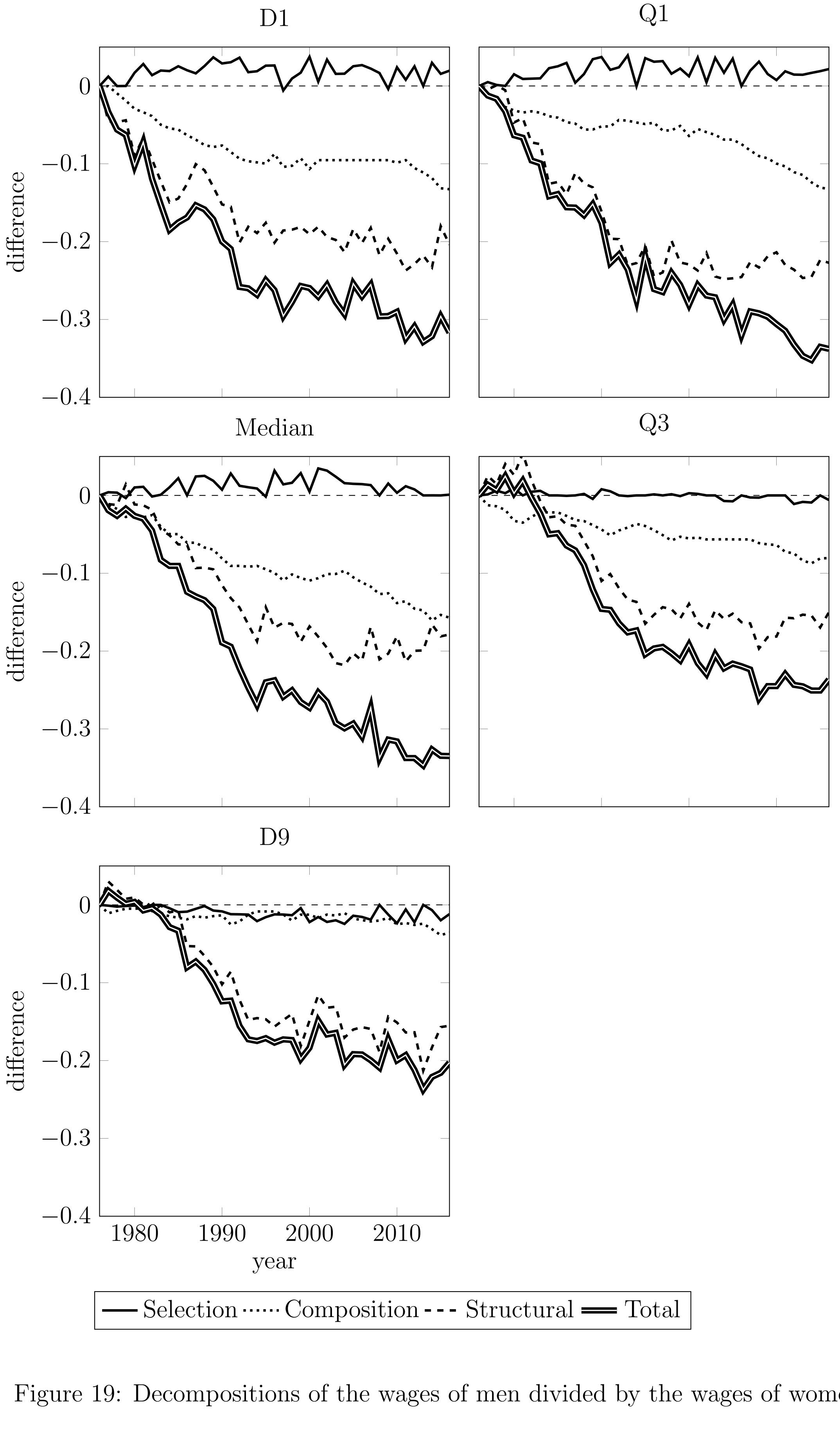

Figure 19: Decompositions of the wages of men divided by the wages of women.

Appendix

We briefly describe the FVV procedure and the estimation of the objects of interest. For further details see FVV. We estimate the control variable as the conditional distribution function of H given Z , via logistic distribution regression. That is:

$$\widehat{V}_i = \Lambda(P(Z_i)^\top \widehat{\pi}(H_i)), \quad i = 1, \ldots, n,$$

where $\Lambda(u) = [1 + \exp(-u)]^{-1}$ is the logistic distribution function, $P(z)$ is a p-dimensional vector of transformations of z involving polynomials, and

$$\widehat{\pi}(h) = \arg\max_{\pi \in \mathbb{R}^p} \sum_{i=1}^{n} \big[\mathbf{1}\{H_i \leq h\} \log \Lambda(P(Z_i)^\top \pi)) + \\ +\mathbf{1}\{H_i > h\} \log\{1 - \Lambda(P(Z_i)^\top \pi)\}\big],$$

for $h \in \mathcal{H}_n$, the empirical support of H.

The estimator of the LASF is $\widehat{\mu}(x, v) = P(x, v)^\top \widehat{\beta}$, where $\widehat{\beta}$ is the least squares estimator and $P(x, v)$ is a p-dimensional vector of transformations of (x, v) involving polynomials and interactions:

$$\widehat{\beta} = \left[\sum_{i=1}^{n} P(X_i, \widehat{V}_i) P(X_i, \widehat{V}_i)^\top\right]^{-1} \sum_{i=1}^{n} P(X_i, \widehat{V}_i)^\top W_i.$$

The estimator of the LDSF is $\widehat{G}(y, x, v) = \Lambda(P(x, v)^\top \widehat{\beta}(w))$, where $\widehat{\beta}(w)$ is the logistic distribution regression estimator:

$$\widehat{\beta}(w) = \arg\max_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \Big[\mathbf{1}\{W_i \leq w\} \log \Lambda(P(X_i, \widehat{V}_i)^\top \beta)) + \\ +\mathbf{1}\{W_i > w\} \log \Lambda(P(X_i, \widehat{V}_i)^\top \beta))\Big].$$

Similarly, the estimator of the LQSF is $\widehat{q}(\tau, x, v) = P(x, v)^\top \widehat{\beta}(\tau)$, where $\widehat{\beta}(\tau)$ is the Koenker and Bassett (1978) quantile regression estimator:

$$\widehat{\beta}(\tau) = \arg\max_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(W_i - P(X_i, \widehat{V}_i)^\top \beta),$$

and $\rho_\tau(u) = [\tau - \mathbf{1}\{u < 0\}]u$ is the "check function".

Estimators of the local derivatives are obtained by taking derivatives of the estimators of the local structural functions. The estimator of the LADF is:

$$\widehat{\delta}(x, v) = \partial_x P(x, v)^\top \widehat{\beta},$$

and the estimator of the LQDF is:

$$\widehat{\delta}_\tau(x, v) = \partial_x P(x, v)^\top \widehat{\beta}(\tau).$$

We obtain estimators of the generic global effects by approximating the integrals over the control variable by averages of the estimated local effects evaluated at the estimated control variable. The estimator of the effect (3) is

$$\widehat{\theta}_{x_0}(x) = \sum_{i=1}^{n} S_i K_i(x_0) \widehat{\theta}(x, \widehat{V}_i) / \sum_{i=1}^{n} S_i K_i(x_0),$$

for $K_i(x_0) = \mathbf{1}\{X_i = x_0\}$ when X is discrete, and $K_i(x_0) = k_h(X_i - x_0)$ when X is continuous, where $k_h(u) = k(u/h)/h$, k is a kernel and h is a bandwidth such as $h \to 0$ as $n \to 0$.

The estimator of the counterfactual distribution is:

$$\widehat{G}^s_{Y_{\langle t|k,r \rangle}}(y) = \frac{1}{n^s_{kr}} \sum_{i=1}^{n} \Lambda(P(X_i, \widehat{V}_i)^\top \widehat{\beta}_t(y)) \mathbf{1}\{\widehat{V}_i > \Lambda(P(Z_i)^\top \widehat{\beta}_r(0))\},$$

where the average is taken over the sample values of $\widehat{V}_i$ and Z_i in group k, $n^s_{kr} = \sum_{i=1}^{n} \mathbf{1}\{\widehat{V}_i > \Lambda(P(Z_i)^\top \widehat{\beta}_r(0))\}$, $\widehat{\beta}_t(y)$ is the distribution regression estimator of step 2 in group t, and $\widehat{\beta}_r$ is the distribution regression estimator of step 1 in group r. Here we are estimating the components $F^s_{Y_t}$ by logistic distribution regression in group t and the component $F^s_{Z_k}$ by the empirical distribution in group k.

We do not have direct empirical analogs for the conditional distribution of the control variable $F^s_{V_r}$ and the support $\mathcal{V}_r(z)$ over all $z \in \mathcal{Z}_k$. We estimate these components using the empirical distribution of $\widehat{V}_i$ in group k, conditional on $\widehat{V}_i > \Lambda(P(Z_i)^\top \widehat{\beta}_r(0))$. This condition selects observations with the values of the explanatory and control variables from group k, which would have been selected with the parameters of the selection equation of group r. This estimation relies on:

$$F_V^s(v \mid Z = z) = \mathbf{1}\{v > F_C(0 \mid Z = z)\}\ v,$$

and the assumption that the ranking of the observations in the conditional distribution of the selection variable is invariant across groups.