Selection and the distribution of female real hourly wages in the United States

IVÁN FERNÁNDEZ-VAL
Department of Economics, Boston University

AICO VAN VUUREN
Department of Economics, Gothenburg University and Department of Economics, Econometrics and Finance, University of Groningen

FRANCIS VELLA
Department of Economics, Georgetown University

FRANCO PERACCHI
Department of Economics and Finance, University of Rome Tor Vergata

We analyze the role of selection bias in generating changes in the observed distribution of female hourly wages in the United States using CPS data for the years 1975 to 2020. We account for selection bias from the employment decision by modeling the distribution of the number of working hours and estimating a nonseparable model of wages. We decompose changes in the wage distribution into composition, structural, and selection effects. Composition effects increased wages at all quantiles while the impact of the structural effects varied by time period and quantile. Changes in the role of selection only appeared at the lower quantiles of the wage distribution. The evidence suggests that there was positive selection in the 1970s, which diminished until the later 1990s. This reduced wages at lower quantiles and increased wage inequality. Post 2000 there appears to be an increase in positive sorting, which reduced the selection effects on wage inequality.

KEYWORDS. Wage inequality, wage decompositions, selection bias.

JEL CLASSIFICATION. C14, I24, J00.

1. INTRODUCTION

The dramatic increase in female wage inequality in the United States since the early 1980s (see, e.g., Katz and Murphy (1992), Katz and Autor (1999), Lee (1999), Autor, Katz,

and Kearney (2008), Acemoglu and Autor (2011), Autor, Manning, and Smith (2016), and Murphy and Topel (2016)) has been accompanied by substantial changes in both female employment rates and the distribution of their annual hours of work. Given the prominence that accounting for the selection bias (see Heckman (1974, 1979)) from employment decisions has played in empirical studies of the determinants of female wages it seems natural to investigate its role in the evolution of wage inequality. This paper examines the sources of changes in the distribution of female hourly real wage rates in the United States from 1975 to 2020 while accounting for movements, and individuals' locations, in the annual hours of work distribution.

The inequality literature allocates wage changes to two sources. The first is the "structural effect", which captures the market value of an individual's characteristics. This includes skill premia, such as the returns to education (see, e.g., Welch (2000) and Murphy and Topel (2016)), cognitive and noncognitive skills (Heckman, Stixrud, and Urzua (2006)), declining minimum wages in real terms (see, e.g., DiNardo, Fortin, and Lemieux (1996), and Lee (1999)), and the increasing use of noncompete clauses in employment contracts (Krueger and Posner (2018)). The second source, referred to as the "composition effect", reflects differences across workers' observed characteristics. These include increases in educational attainment. Earlier papers (see, e.g., Angrist, Chernozhukov, and Fernández-Val (2006) and Chernozhukov, Fernández-Val, and Melly (2013)) have estimated these two effects under general conditions. However, as they focus on male wages they have ignored the role of selection.

Understanding the role of selection in the female wage inequality context is important. First, as the impact of selection is frequently interpreted as reflecting sorting patterns, it is valuable from a policy perspective to understand how worker productivity has changed as an increasing proportion of women have entered the labor market. Second, the importance of accounting for selection bias in estimating the determinants of female wages suggests that an evaluation of the role of structural and composition effects requires an appropriate treatment of selection. Third, assessing the impact of selection on wages and inequality is particularly relevant when the composition of the working and nonworking populations have evolved as drastically as has occurred in our sample period. Finally, understanding the impact of selection bias may provide policymakers with guidance as to which measures may be taken to reduce wage inequality.

Three important papers have investigated the role of selection in the United States over a period of increasing female wage inequality.[1] Mulligan and Rubinstein (2008), hereafter MR, correct for selection in the female mean wage in the United States for the years 1975–1999 and argue that the sharp increase of female wages partially reflected that the selected population of working females became increasingly more productive in terms of unobservables. They also find that the pattern of sorting turned from negative to positive in the early 1990s. Evaluating the contribution of selection bias on the mean wage is straightforward in additive models as the selection component, under some assumptions, can be separated. The pattern of sorting is inferred from the coefficient for

---

[1]Fernández-Val, van Vuuren, Vella, and Peracchi (2022) study the impact of annual hours worked on inequality but their focus is on annual earnings and not wages.

the selection correction. The nonseparable model required for estimating wage distributions has greater difficulties in isolating the selection component. Maasoumi and Wang (2019), hereafter MW, employ the copula based estimator for quantile selection models of Arellano and Bonhomme (2017), hereafter AB. AB and MW define the selection effect as the difference between the observed wage distribution and the counterfactual wage distribution simulated via their models' estimates assuming 100% participation. MW provide a similar conclusion regarding the pattern of sorting as MR for the overlapping years in their studies. Blau, Kahn, Boboshko, and Comey (2021) follow Olivetti and Petrongolo (2008) who use a "selection on unobservables" approach to impute wages for nonworkers based on propensity scores for employment. They also compute the predicted wage distribution assuming 100% participation. In contrast to MR and MW, they find a more modest role for selection and that sorting did not change sign over the sample period. We define the selection effect as the difference in the observed wage distribution and the wage distribution that would result under the participation process associated with the year with the lowest participation rate. We find that the direction of sorting did not change during the years considered by MR.

We address several methodological and empirical issues regarding selection in the female wage inequality context. Our methodological contributions are the following. First, we extend the Fernández-Val, van Vuuren, and Vella (2021), hereafter FVV, estimator for nonseparable models with censored selection rules. The FVV estimator incorporates the number of working hours rather than the binary work decision as the selection variable and here we allow for different censoring points conditional on the individual's characteristics. This variation across censoring points captures differences in "fixed working costs" (see, e.g., Cogan (1981)). Second, we provide a procedure for decomposing changes in the wage distribution into structural, composition, and selection effects in a nonseparable model, which allows for selection from the choice of annual working hours. We contrast our decomposition approach with the corresponding exercise based on the Heckman (1979) selection model (HSM). Third, we extend our estimator to allow selection into annual hours to reflect two separate selection mechanisms. Namely, the choices of annual weeks and weekly hours. Fourth, we provide an estimator motivated by the ordered treatment model of Heckman and Vytlacil (2007), which allows for bunching in annual hours or annual weeks and apply it via our decomposition method.

Our following empirical contributions feature results based on the two most commonly employed Current Population Survey (CPS) data sets. First, unlike MR and MW who analyze wages for full-time full-year (FTFY) workers, we obtain a fuller picture of the evolution of the wage distribution by including all workers and accounting for selection from the hours of work decision. Second, we confirm previous findings, restricted to FTFY workers, regarding movements in the wage distribution. Female wage growth at lower quantiles is modest although the median wage has grown steadily. Gains at the upper quantiles are large and have produced an increase in female wage inequality. Finally, we provide new evidence regarding the role of selection. Changes in selection are especially important at the lower end of the wage distribution and have generally decreased

wage growth and increased wage inequality. Although we are able to reproduce the estimated sorting pattern as MR and MW, we illustrate that this reflects the employed identification assumptions. We show that exploiting the variation in hours worked as a form of identification produces results consistent with positive sorting for the whole sample period.

An important empirical result relates to the pattern of sorting and its implication for the impact of selection on wages and inequality. We find clear evidence of positive sorting in the mid 1970s. The period 1975 to 2000 experienced a shift in the distribution of female annual hours of work, accompanied by a reduction in the level of positive sorting. These two forces decreased wages at lower quantiles and increased wage inequality. For the remainder of our sample period, there appears to be a return to higher levels of positive sorting and a decrease in the impact of selection on wage inequality.

The rest of the paper is organized as follows. The next section discusses the data. Section 3 describes our empirical model and defines our decomposition exercise. It also provides alternative estimators employing ordered or multiple censored selection rules. This section concludes with a comparison of our decomposition approach with that associated with the HSM. Section 4 outlines how the methodology is implemented and Section 5 presents the empirical results. Section 6 reconciles the difference between our results with those of MR, while Section 7 investigates the impact of the changes in selection in wage inequality. Section 8 offers some concluding comments. Supplementary work is provided in the Online Supplementary Material in the Appendix (Fernández-Val, van Vuuren, Vella, and Peracchi (2023)). References to this Appendix are made in the text.

2. Data

We employ the two most commonly analyzed microlevel data sets, the Annual Social and Economic Supplement (ASEC) and the Merged Outgoing Rotation Groups (MORG), from the CPS. Appendix A of Lemieux (2006) provides a comparison of the two data sets. We employ both to contrast results and to allow comparisons with earlier studies.

2.1 *Annual social and economic supplement*

We employ the ASEC for the 46 survey years from 1976 to 2021 reporting annual earnings for the previous calendar year.[2] Unless otherwise stated, we refer to the year for which the data are collected and not that of the survey. The 1976 survey is the first for which information on weeks worked and usual hours of work per week last year are available. To avoid issues related to retirement and ongoing educational investment, we restrict attention to those aged 24–65 years in the survey year. This produces an overall sample of 2,219,820 females. The annual sample sizes range from 33,924 in 1976 to 59,622 in 2001.

Annual hours worked are defined as the product of weeks worked and usual weekly hours of work last year. Those reporting zero hours usually respond that they are not

---

[2]The data were obtained via the IPUMS-CPS website maintained by the Minnesota Population Center (Flood et al. (2015)).

in the labor force in the week of the March survey. We define hourly wages as the ratio of reported annual labor earnings in the year before the survey, converted to constant 2019 prices using the consumer price index for all urban consumers, and annual hours worked. Hourly wages are unavailable for those not in the labor force. For the self-employed, unpaid family workers and the Armed Forces annual earnings or annual hours tend to be poorly measured and we exclude these groups from our sample. This results in a deletion of 5.4, 0.4 and 0.07% of the observations for the self-employed, unpaid family workers, and the Armed Forces, respectively. The figures for self-employed and the armed forces have trended upwards while those for family workers have trended downwards over the sample period. These groups do not show any cyclical variation. The only exception is the number of self employed during the Great Recession, which compared to the total employed, dropped considerably. We use observations with imputed wages for their values of working hours but do not use them in the wage sample. The restriction to civilian dependent employees with positive hourly wages and people out of the labor force last year results in a sample of 2,055,063 females. The subsample of civilian dependent employees with positive hourly wages comprises 1,190,928 observations. A benefit of the ASEC is its extensive family background variables.

2.2 *Merged outgoing rotation groups*

We use the years 1979 to 2019 for the MORG using the CEPR extracts. The MORG contains information on hourly wages in the survey week for those paid by the hour and on weekly earnings from the primary job during the survey week for those not paid by the hour. Lemieux (2006) and Autor, Katz, and Kearney (2008) argue that the ORG CPS data are preferable as they provide a point-in-time wage measure and workers paid by the hour, more than half of the U.S. workforce, may better recall their hourly wages. Based on comparable data restrictions as the ASEC, the MORG results in 4,298,682 observations with the highest numbers of observations in 1980 (121,786) and the lowest in 2019 (91,647). The subsample of civilian dependent employees working in the reference week is 2,219,820 observations. This low figure, relative to the ASEC, is expected as employees who did not work in the reference week may have worked in another week. Family background variables are only available since 1984. This restricts the family background characteristics to family size.

2.3 *Descriptive statistics*

Figure 1 presents female real hourly wage rates at selected quantiles for both the ASEC and MORG. The shaded areas in Figure 1 indicate recessionary periods as dated by the National Bureau of Economic Research. Quartiles are denoted by Q and deciles by D. Figure 1 confirms two observations made by Lemieux (2006). First, median wages for the ASEC and MORG are very similar. The median wage in the ASEC, despite occasional dips, increases by 32.5 percentage points over the period 1975–2020 while the MORG shows very similar patterns with growth of 25.9% for 1979–2019. Second, the ASEC features more wage dispersion with relatively lower wages at quantiles below Q2, and higher
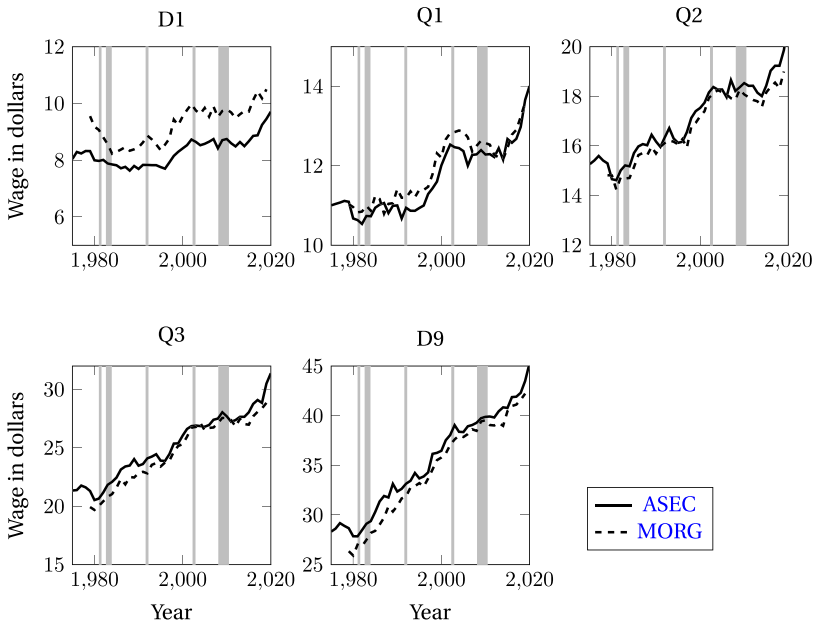
FIGURE 1. Real hourly wage at different quantiles (measured in 2019 dollars).

wages at quantiles above Q2. The profiles at D1 and Q1 are similar to that at Q2 with increases of 18.8% and 24.1%, respectively. At Q3 and D9, there has been strong and steady growth since 1980 with an increasing gap between each and the median wage. Q3 increases by 40.2% and D9 by 48.4%. The MORG shows very similar patterns with growth of 25.9% for the median wage for 1979–2019, and increases of 22.5%, 37.1%, and 45.8% at D1, Q3, and D9. The corresponding increase in the ASEC median wage for the same time period is 26.8%. The only remarkable difference across the figures are the wages at D1, which are clearly higher for the MORG than for the ASEC. This, as noted by Lemieux, may reflect measurement error. However, measurement error does not explain why the difference is greater at D1 than D9. Another cause, as we explain below, may be the relatively lower employment rate in the MORG. This implies that the MORG D1 is higher in the population distribution than the ASEC D1. The difference between the data sets decreases for the MORG in 1979–1981 with a corresponding smaller decrease for the ASEC. The ASEC and the MORG then show similar growth with the ASEC wage consistently below the MORG. As noted by Lemieux (2006), the period 1979 to 1984 displays a sharp increase in the residual variance in the MORG not found for the ASEC.

Figure 2 confirms the widening wage gap with an increase in the interquartile ratio for our sample period. For the ASEC there is an increase of 49.4%. For the shorter period the MORG data features an increase of 45.0% with the corresponding increase in the ASEC for that period of 44.3%. Both data sets confirm the drastic increases in female wage inequality. The pattern associated with the interdecile ratio is somewhat more complicated. Overall, the MORG has the larger increase but the difference reflects wage movements for the period 1979 to 1984. The increase in the interdecile ratio post
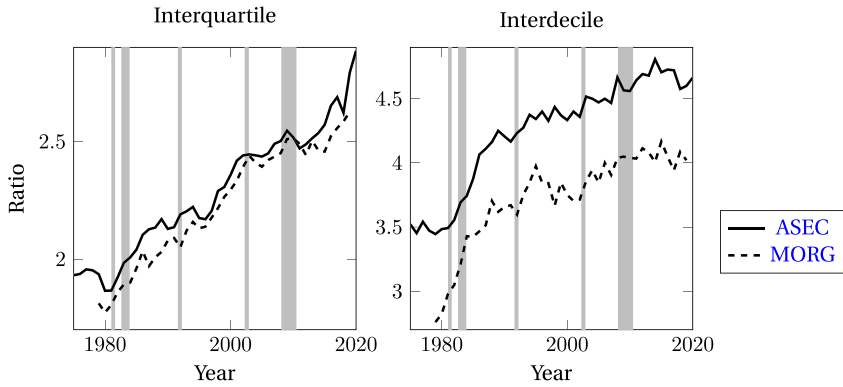
FIGURE 2. Measures of inequality.

1984 is relatively lower for the MORG. This is consistent with Lemieux (2006) who notes that the ASEC not only has higher wage dispersion but also increases faster over time.

Figure 3 presents the female employment rates and average working weeks and hours of those employed. As explained above, the higher employment rate for ASEC is expected. The employment rates increased drastically during the last decades of the previous century before decreasing slightly in the 2000s and then slightly rebounding in the most recent years. The decreased employment since 2000 resulted in the U.S. employment rate at the end of the Great Recession returning to the 1980 level (see Beaudry, Green, and Sand (2016)). The average number of working hours per week increased over the sample period from 35.3 to 38.3 for the ASEC and from 35.3 to 37.8 for the MORG. Hours are highly cyclical. There is not a large difference between the two samples although the increase in working hours has been slightly smaller in the 1990s for the MORG. The number of weeks also increased steeply from 42.0 to 46.2. The drop in 2020 during the COVID-19 pandemic is noteworthy.
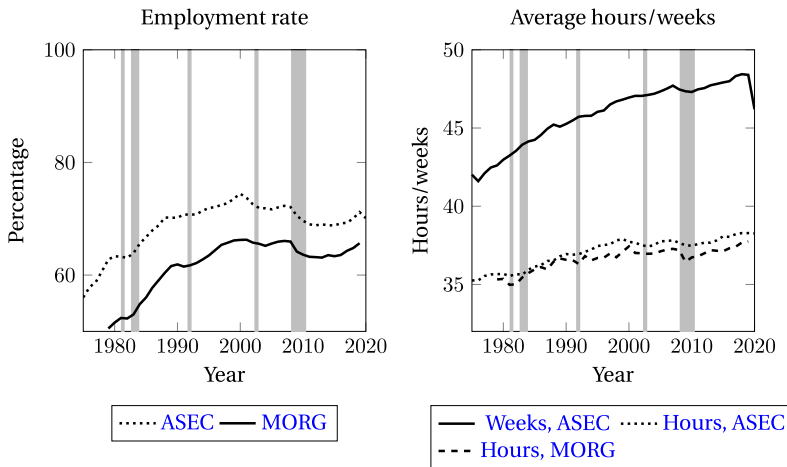

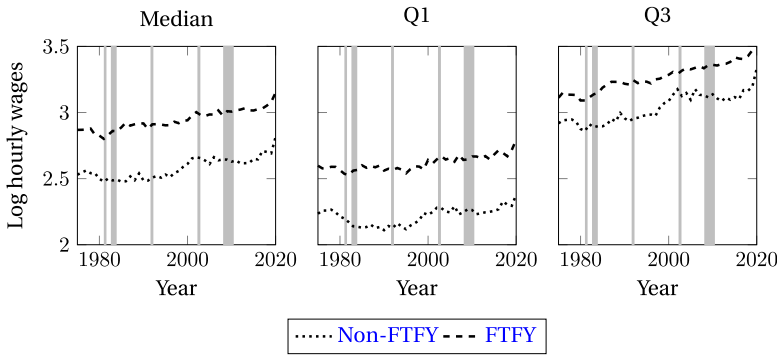FIGURE 3. Employment rates and hours/weeks worked of employees.

FIGURE 4. Percentiles of real log hourly wages by full-time full-year (FTFY) employment status.

Figure 4 reveals that the wage levels of FTFY workers, defined as working 52 weeks per year for at least 40 hours, are somewhat higher than for all workers. Their fraction increased from 50 to 71%. The difference in wages diminishes at higher quantiles and, though not reported here, becomes negligible at D9. Contrasting full-year (FY) and full-time (FT) wages against non-FY and non-FT wages provides a similar finding. The differences at the bottom of the distribution in comparison to the top, may reflect that the choice of working FT and/or FY should be incorporated in the selection model. Moreover, as non-FTFY (or non-FT/non-FY) workers are potentially more vulnerable to interruptions it seems inappropriate to exclude them in an evaluation of wage inequality. The wage differences between these two working categories might also reflect selection effects. The failure to include those who do not work FTFY means that the selection effects in earlier studies may reflect movement from the non-FTFY to FTFY, rather than from nonemployment to FTFY.

3. ECONOMETRIC ANALYSIS

3.1 *Model with censored selection*

We consider a version of the HSM where the censoring rule for the selection process incorporates the information on annual hours worked rather than the binary employment/nonemployment decision. The model has the form:

$$Y = g(X, E), \quad \text{if } D = 1, \tag{1}$$

$$H = h(Z, V), \quad \text{if } D = 1, \tag{2}$$

$$D = 1\{h(Z, V) > \mu(Z)\}, \tag{3}$$

where Y is the logarithm of hourly wages, H is annual hours worked, D is a selection or employment indicator equal to 1 if H and Y are observed or equivalently if H is not censored at $\mu(Z)$, X and Z are vectors of observable conditioning variables, g, h, and μ are unknown functions, and E and V are respectively a vector and a scalar of potentially dependent unobservable variables with cumulative distribution functions (CDFs) F_E and F_V. We assume that X is a subset of Z, not necessarily strict, that is, $X \subseteq Z$.

We refer to equation (3) as the selection rule. It corresponds to censored selection with an unobserved censoring point. That is, we observe the censoring status, D, but not the censoring point, $\mu(Z)$. Equations (2)–(3) can be considered a reduced-form representation for hours worked. The model is a nonparametric and nonseparable version of the Tobit type-3 model considered by FVV, extended to incorporate an unknown censoring threshold, which is a function of Z. This threshold is motivated by fixed labor costs measured in terms of hours. Individuals only work if the desired number of hours exceeds a minimum number given by $\mu(Z)$. Cogan (1981) shows that fixed labor costs reduce the number of individuals working very few hours. We allow the fixed labor costs to vary by individual and household characteristics.

Let $\perp\!\!\!\perp$ denote stochastic independence. We assume the following.

Assumption 1 (Control Function). $(E, V) \perp\!\!\!\perp Z, V$ *is a continuously distributed random variable with strictly increasing CDF on the support of V, and $v \mapsto h(Z, v)$ is strictly increasing a.s.*

Without loss of generality, we can normalize V to be standard uniformly distributed. The potential dependence between E and V implies $Z's$ independence of E in the entire population does not exclude dependence in the selected population with $D = 1$. FVV showed that Z is independent of E conditional on V and $D = 1$ when $\mu(Z) = \mu$. Let $H^* = h(Z, V)$. Lemma 1 extends FVV's result to our model.

Lemma 1 (Existence of Control Function). *Under the model in (1)–(3) and Assumption 1:*

$$E \perp\!\!\!\perp Z \mid V, D = 1.$$

Moreover, $V = F_{H^*\mid Z}(H \mid Z)$ for $H^* = h(Z, V)$, and $F_{H^*\mid Z}$ is identified by

$$F_{H^*\mid Z}(h \mid z) = 1 - \pi(z)\big[1 - F_{H\mid Z,D}(h \mid z, 1)\big],$$

where $\pi(z) = \mathbb{P}(D = 1 \mid Z = z)$ is the propensity score of selection.

The proof of the first statement follows from the same argument as in Lemma 1 of FVV. Thus, conditioning on (Z, V) makes $D = 1$ or $h(Z, V) > \mu(Z)$ deterministic. The assumption that Z is independent of (E, V) then proves the result implying that V is an appropriate control function.[3] The result $V = F_{H^*\mid Z}(H \mid Z)$ follows directly from the assumption that h is strictly increasing in its second argument and the normalization on the distribution of V. Identification of $F_{H^*\mid Z}$ follows from Buchinsky and Hahn (1995). We propose an estimator of $F_{H^*\mid Z}$ based on distribution regression in the Online Appendix A. This estimator is an alternative to the estimators of Buchinsky and Hahn (1998) and Chernozhukov and Hong (2002), which are based on quantile regression.

We denote the support of random variables and vectors by calligraphic letters while lower case letters in parentheses indicate that the support is conditional on a stochastic

---

[3]This result is closely related to that of Imbens and Newey (2009), who consider estimation and identification of a nonseparable model with a single continuous endogenous explanatory variable.

vector taking a particular value; for example, $\mathcal{Z}(x)$ is the support of $Z \mid X = x$. We define the Local Average Structural Function (LASF) and Local Distribution Structural Function (LDSF) as

$$m(x, v) := \mathbb{E}\big[g(x, E) \mid V = v\big], \qquad G(y, x, v) := \mathbb{P}\big[g(x, E) \leq y \mid V = v\big]. \qquad (4)$$

They represent the mean and distribution of Y if all individuals with control function equal to v had observable characteristics equal to x. An argument similar to FVV shows that

$$m(x, v) = \mathbb{E}[Y \mid V = v, X = x, D = 1], \qquad G(y, x, v) = \mathbb{P}[Y \leq y \mid V = v, X = x, D = 1],$$

provided $(x, v) \in \mathcal{XV}^*$, where

$$\mathcal{XV}^* = \big\{(x, v) \in \mathcal{X} \times \mathcal{V} \mid \exists z \in \mathcal{Z}(x) : h(z, v) > \mu(z)\big\}.$$

This set, referred to as the identification set by FVV, is identical to the support of (X, V) among the selected population. Lemma 1 implies that the LASF and LDSF equal the mean and distribution of the observed Y conditional on (X, V) and that it is identified. This follows directly from $(E, V) \perp\!\!\!\perp Z$ and that $(x, v) \in \mathcal{XV}^*$ implies the ability to find a (z, v) combination for which $h(z, v) > \mu(z)$. We refer to FVV for a discussion on how the size of the identified set depends on the availability of exclusion restrictions on Z with respect to X.

There are different candidates for H in (2). As the ASEC provides both usual hours worked per week and annual hours, calculated as the product of weeks worked last year and the usual number of hours worked per week, we employ several alternatives. Although the usual hours per week may be the variable in the ASEC that is closest to the hours decision in labor supply models (Killingsworth (1983)), it may also reflect whether the job has pre-set hours. Therefore, we employ the annual measure, which incorporates the weeks decision. As the extensive margin may capture whether an individual has worked a positive number of hours in the past year, we also investigate the use of the number of worked weeks. A theoretical motivation for this measure follows from search models in which the offered wages depend positively on the job offer arrival rate and negatively on the separation rate (Burdett and Mortensen (1998)). As these rates also determine the number of weeks worked, it implies a relationship between weeks worked and wages. The appropriate censoring variable in the MORG is the number of hours worked in the reference week. Note that this variable solves some of the problems mentioned above.

3.2 *Counterfactual distributions*

We consider counterfactual CDFs constructed by integrating the LDSF with respect to different joint distributions of the conditioning variables and control function.[4] Counterfactual CDFs enable the construction of wage decompositions and facilitate counterfactual analyses similar to those in DiNardo, Fortin, and Lemieux (1996), Ñopo (2008),

---

[4]Counterfactual means can be constructed similarly using the LASF in place of the LDSF; see Section 3.5.

Fortin, Lemieux, and Firpo (2011), and Chernozhukov, Fernández-Val, and Melly (2013). We focus on CDFs for the selected population. To simplify notation, the superscript s denotes that we condition on $D = 1$. The decompositions are based on the following representation of the CDF of the observed Y:[5]

$$\mathbb{P}[Y \leq y \mid D = 1] =: G_Y^s(y) = \int G(y, x, v) \, dF_{Z,V}^s(z, v),$$

where

$$F_{Z,V}^s(z, v) = \frac{1\{h(z, v) > \mu(z)\} F_{Z,V}(z, v)}{\int 1\{h(z, v) > \mu(z)\} \, dF_{Z,V}(z, v)},$$

denotes the joint CDF of (Z, V) in the selected population and $F_{Z,V}$ denotes the joint CDF of Z and V in the entire population.

The counterfactual CDFs are constructed by combining the CDFs G and $F_{Z,V}$ with the selection rule (3) for different groups, each group corresponding to a different time period or a subpopulation defined by certain characteristics. Specifically, let G_t be the LDSF in group t, F_{Z_k,V_k} be the joint CDF of Z and V in group k, and let $1\{h_r(z, v) > \mu_r(z)\}$ be the selection rule in group r. The counterfactual CDF of Y when G is as in group t, $F_{Z,V}$ is as in group k, and the selection rule is as in group r is defined as

$$G_{Y_{\langle t,k,r\rangle}}^s(y) = \frac{\int G_t(y, x, v) \, 1\{h_r(z, v) > \mu_r(z)\} \, dF_{Z_k,V_k}(z, v)}{\int 1\{h_r(z, v) > \mu_r(z)\} \, dF_{Z_k,V_k}(z, v)}, \tag{5}$$

provided that the integrals are well-defined. Since the mapping $v \mapsto h(z, v)$ is strictly monotonic, the condition $h_r(z, v) > \mu_r(z)$ in (5) is equivalent to the condition:

$$v > F_{H_r^* \mid Z}\big(\mu_r(z) \mid z\big) = 1 - \pi_r(z), \tag{6}$$

where $F_{H_r^* \mid Z}(\mu_r(z) \mid z)$ is the probability of working less hours than the censoring point conditional on $Z = z$ in group r and $\pi_r(z)$ is the propensity of working in that group. Given $G_{Y_{\langle t,k,r\rangle}}^s(y)$, the corresponding counterfactual quantile function (QF) is

$$q_{Y_{\langle t,k,r\rangle}}^s(\tau) = \inf\{y \in \mathbb{R}: G_{Y_{\langle t,k,r\rangle}}^s(y) \geq \tau\}, \quad 0 \leq \tau \leq 1. \tag{7}$$

Under these definitions, the observed CDF and QF of Y for the selected population in group t are $G_{Y_{\langle t,t,t\rangle}}^s$ and $q_{Y_{\langle t,t,t\rangle}}^s$, respectively.

Nonparametric identification of (5) and (7) depends on whether the integrals in (5) are well-defined. They are when two conditions are met. First, if $\mathcal{Z}_k \subseteq \mathcal{Z}_r$, then π_r is identified over all z combinations in the integral. Second, when $(\mathcal{XV}_k \cap \mathcal{XV}_r^*) \subseteq \mathcal{XV}_t$, then the LDSF is identified for all combinations of z on which we integrate. Here, $\mathcal{XV}_r^*$ denotes the support of (X, V) for the selected population in group r. The identification conditions simplify when we consider two years for q, r, and t, such as 0 and 1, which is

---

[5]We refer to FVV for details.

relevant for the decompositions. For example, we need $\mathcal{XV}_1 \cap \mathcal{XV}_0^* \subseteq \mathcal{XV}_1^*$ and $\mathcal{Z}_1 \subseteq \mathcal{Z}_0$ to identify $G_{\langle 1,1,0 \rangle}$ and $\mathcal{XV}_0^* \subseteq \mathcal{XV}_1^*$ to identify $G_{\langle 1,0,0 \rangle}$. A sufficient condition for $\mathcal{XV}_1 \cap \mathcal{XV}_0^* \subseteq \mathcal{XV}_1^*$ and $\mathcal{XV}_0^* \subseteq \mathcal{XV}_1^*$ is that the employment rates in year 0, conditional on X, are lower than those in year 1.

Using (7), we decompose the difference in the observed QF of Y for the selected population between any two groups, say group 1 and group 0, as[6]

$$q^s_{Y\langle 1,1,1 \rangle} - q^s_{Y\langle 0,0,0 \rangle} = \underbrace{\left[q^s_{Y\langle 1,1,1 \rangle} - q^s_{Y\langle 1,1,0 \rangle} \right]}_{[1]} + \underbrace{\left[q^s_{Y\langle 1,1,0 \rangle} - q^s_{Y\langle 1,0,0 \rangle} \right]}_{[2]} + \underbrace{\left[q^s_{Y\langle 1,0,0 \rangle} - q^s_{Y\langle 0,0,0 \rangle} \right]}_{[3]}, \quad (8)$$

where [1] is a selection effect that captures changes in the selection rule given the joint distribution of Z and V, [2] is a composition effect that reflects changes in the joint distribution of Z and V, and [3] is a structural effect that reflects changes in the conditional distribution of Y given Z and V. These effects are relative to the base year. We stress that this definition of the selection effect differs from the standard definition. This is discussed in Section 3.5.

3.3 *Double selection mechanism*

The model can be extended to a multiple censored selection mechanism operating through both weeks and hours. The model has the form:

$$Y = g(X, E), \quad \text{if } h(Z, V_H) > \mu_H(Z) \text{ and } w(Z, V_W) > \mu_W(Z), \quad (9)$$

$$H = h(Z, V_H), \quad \text{if } h(Z, V_H) > \mu_H(Z), \quad (10)$$

$$W = w(Z, V_W), \quad \text{if } w(Z, V_W) > \mu_W(Z), \quad (11)$$

where the unobserved thresholds $\mu_H(Z)$ and $\mu_W(Z)$ are functions of the individual's characteristics.[7] The distributions of $H^* := h(Z, V_H)$ and $W^* := w(Z, V_W)$ conditional on Z, denoted V_H and V_W, respectively, are each required as control functions. The analysis of this model is similar to that above. However, it is necessary to employ both control functions to, for example, calculate the LDSF, that is,

$$G(y, x, v_H, v_W) = \mathbb{P}\big(g(x, E) \leq y \,|\, V_H = v_H, V_W = v_W\big).$$

The identification conditions change to accommodate that the support condition is defined over two control functions. Using the same notation as Section 3.2, the support requirements for the counterfactuals are $\mathcal{ZV}_H\mathcal{V}_{W,k} \subseteq \mathcal{ZV}_H\mathcal{V}_{W,r}$ and $(\mathcal{XV}_H\mathcal{V}_{W,k} \cap \mathcal{XV}_H\mathcal{V}_{W,r}^*) \subseteq \mathcal{XV}_H\mathcal{V}_{W,t}$. We acknowledge that there are circumstances under which this model will collapse to the single censoring mechanism case. However, as these are somewhat obvious we do not detail them here.

---

[6]Note that alternative orders are possible. We investigated the impact of changes in the order on our empirical results and these were minor. Details are available upon request.

[7]A further complication could be introduced by assuming that either the equation (10) or (11) is the sole source of selection.

3.4 *Model with ordered selection*

The models above employ control functions, which assume that the selection variable is continuous. However, both the numbers of weeks and hours worked feature bunching at specific values (e.g., 40 hours and 52 weeks). The following model with ordered selection incorporates bunching:

$$Y = g(X, E) \quad \text{if } H \geq 1,$$

$$H = \begin{cases} 0 & \text{if } V \leq \mu_0(Z), \\ 1 & \text{if } \mu_0(Z) < V \leq \mu_1(Z), \\ \vdots & \\ K & \text{if } V > \mu_{K-1}(Z), \end{cases}$$

where the variables have similar interpretations as above. The main difference between this model and those above is that it allows a discrete distribution of H at the expense of requiring separability in the selection process. We assume $Z \perp\!\!\!\perp (E, V)$ and V follows a standard uniform distribution. This model is related to the ordered choice model of Heckman and Vytlacil (2007, p. 4980), but unlike their model $g(X, E)$ does not depend on H. It can also be interpreted as an extension of Newey (2007) to multiple ordered outcomes.

We define the identification set as

$$\mathcal{XP}_K := \left\{ (x, p) \in \mathcal{X} \times [0, 1) \,|\, \exists (h, z) \in \mathcal{HZ}(x) : \mu_{h'}(z) = p, \, h' \leq h, \, h > 0 \right\}.$$

This set collects (x, p) combinations in the selected sample (i.e., $H = h > 0$) for which there is a (h, z) combination in $\mathcal{HZ}(x)$ such that $\mu_{h'}(Z) = p$ for the propensity score of a value of H smaller or equal than the observed value h. For example, if $H = 3$, this restriction is satisfied when $\mu_{h'}(z) = p$, for some $h' \in \{0, 1, 2, 3\}$. We define the LDSF as in (4). We prove the following lemma in Online Appendix B.

LEMMA 2. *Suppose that* $(x, p) \in \mathcal{XP}_K$ *and* $\mathcal{Z}(x, h) = \mathcal{Z}(x)$ *for all* $h > 0$. *Then*

$$\frac{1}{1 - p} \int_p^1 G(y, x, v) \, dv = \mathbb{P}(Y \leq y \,|\, X = x, H > h, \mu_h(Z) = p),$$

and the probability in the RHS is identified.

Lemma 2 implies $(x, p) \in \mathcal{XP}_K$ is also a sufficient condition for identification as (see also Heckman and Vytlacil (2007)):

$$G(y, x, p) = -\frac{\partial}{\partial p} \int_p^1 G(y, x, v) \, dv = \frac{\partial}{\partial p} \{ (1 - p) \mathbb{P}(Y \leq y \,|\, X = x, H > h, \mu_h(Z) = p) \}.$$

We need additional assumptions to obtain counterfactual distributions. In the models with continuous censoring we hold the value of the control function constant and

change the lowest value at which the individual is participating (see (6)). We cannot follow the same strategy here as V is not point identified. However, from the values of H and Z we know that the value of V is between $\mu_{H-1}(Z)$ and $\mu_H(Z)$. This implies that individuals with $H = 1$ have the lowest values of V. Therefore, if we increase $\mu_0(Z)$, while leaving V unchanged, some individuals with $H = 1$ would no longer participate although we do not know who. Hence, we integrate over the distribution of V and change the range of integration accordingly. We show in the Online Appendix C that

$$G_Y^s(y) = \frac{\sum_{h=1}^{K} \int_{\mathcal{Z}} \int_{\mu_{h-1}(z)}^{\mu_h(z)} G(y, x, v)\, dv\, dF_Z(z)}{\int_{\mathcal{Z}} \left(1 - \mu_0(z)\right) dF_Z(z)}, \tag{12}$$

where $\mu_K(z) := 1$ for any z. This equation is comparable to equation (7.2) of Heckman and Vytlacil (2007). Based on (12), the counterfactual distribution when G is as in group t, F_Z is as in group k, and the selection rule is as in group r is

$$G_{Y_{\langle t,k,r \rangle}}^s(y) = \frac{\int_{\mathcal{Z}^k} \int_{\mu_0^r(z)}^{\mu_1^k(z)} G_t(y, x, v)\, dv\, dF_{Z^k}(z)}{\int_{\mathcal{Z}^k} \left(1 - \mu_0^r(z)\right) dF_{Z^k}(z)}$$

$$+ \frac{\sum_{h=2}^{K} \int_{\mathcal{Z}^k} \int_{\mu_{h-1}^k(z)}^{\mu_h^k(z)} G_t(y, x, v)\, dv\, dF_{Z^k}(z)}{\int_{\mathcal{Z}^k} \left(1 - \mu_0^r(z)\right) dF_{Z^k}(z)}. \tag{13}$$

The decompositions are identical to (8). The identification restrictions are related to the integrals in (13). The integral in the numerator of the second line of (13) can be written as

$$\int_{\mathcal{Z}^k} \left[\int_{\mu_{h-1}^k(z)}^1 G_t(y, x, v)\, dv - \int_{\mu_h^k(z)}^1 G_t(y, x, v)\, dv \right] dF_{Z^k}(z).$$

For both of these terms to be identified for any h, we need that $\mathcal{XP}_K^k \subseteq \mathcal{XP}_K^t$. For the identification of the integral in the numerator of the first line of (13), a similar argument gives $(\mathcal{XP}_K^r \cap \mathcal{XP}_K^k) \subseteq \mathcal{XP}_K^t$. We also need that $\mathcal{Z}^k \subseteq \mathcal{Z}^r$ otherwise the integrals in the denominators are not identified. The identification restrictions imply that, for example, to identify $G_{Y_{\langle 1,1,0 \rangle}}^s$, one needs that $\mathcal{XP}_K^0 \subseteq \mathcal{XP}_K^1$ and $\mathcal{Z}^1 \subseteq \mathcal{Z}^0$. The interpretation of these restrictions not only depends on the employment rates between year 0 and 1 but also on whether the propensity scores in year 1 overlap those of year 0.

Despite these requirements, there is a benefit of using an ordered rather than dichotomous selection rule. In the latter case, the restriction for $G_{Y_{\langle 1,1,0 \rangle}}^s$ would have been that the support of the propensity scores of employment for year 1, $\mu_0^1(Z)$, should overlap with those of year 0, $\mu_0^0(Z)$. For ordered selection, it is only necessary that one of

the propensity scores in year 1, that is, $\mu_h^1(Z)$, $h = 1, \dots, K$, overlaps with the propensity score of employment for year 0, that is, $\mu_0^0(Z)$.

3.5 *Comparison with the Heckman selection model*

The decomposition (8) yields different effects to those derived by MR for a parametric version of the HSM. The MR selection effect excludes one component that we attribute to the selection effect and includes another that we assign to the composition effect. Two other components are sorting effects that cannot be separately identified from the structural effect in nonseparable models. A common measure of a selection effect is the difference between the average wage for the selected population and that for the entire population. The change in the selection effect is how this difference varies over time. We define the change in the selection effect as the difference between the average wage in the selected population with that which would result under a more restrictive selection rule.

To illustrate the difference with MR, suppose that the population model in period t is the following parametric version of the HSM:

$$Y_t = \alpha_t^T X_t + E_t, \quad \text{if } H_t > 0, \tag{14}$$

$$H_t = \max\{\gamma_t^T Z_t + V_t, 0\}, \tag{15}$$

where the first element of X_t is the constant term, and E_t and V_t are distributed independently of (X_t, Z_t) as bivariate normal with zero means, variances $\sigma_{E_t}^2$ and $\sigma_{V_t}^2$, and correlation coefficient ρ_t.[8] The counterfactual mean of Y for the selected population when the LASF is as in group t, $F_{Z,V}$ is as in group k, and the selection rule is as in group r, is

$$m_{Y_{\langle t,k,r \rangle}}^s = \frac{\int m_t(x,v)\, 1\{v > \Phi(-\gamma_r^T z)\}\, dF_{Z_k,V_k}(z,v)}{\int 1\{v > \Phi(-\gamma_r^T z)\}\, dF_{Z_k,V_k}(z,v)}, \tag{16}$$

where

$$m_t(x, v) = \alpha_t^T x + \rho_t \sigma_{E_t} \Phi^{-1}(v),$$

denotes the LASF in group t. The observed mean of Y_t in the selected population, integrating over Z_t, is

$$m_{Y_{\langle t,t,t \rangle}}^s = \alpha_t^T \mathbb{E}[X_t \,|\, H_t > 0] + \rho_t \sigma_{E_t} \mathbb{E}[\lambda(\gamma_t^T Z_t / \sigma_{V_t}) \,|\, H_t > 0],$$

where λ denotes the inverse Mills ratio. We decompose the difference $m_{Y_{\langle 1,1,1 \rangle}}^s - m_{Y_{\langle 0,0,0 \rangle}}^s$ between two time periods, $t = 0$ and $t = 1$, into selection, composition, and structural effects.

---

[8]We present the selection rule in this censored form in order to employ our approach although our discussion of the HSM, which follows, is based on the binary rule that only $1(H_t > 0)$ is observed.

MR define the selection effect as

$$\rho_1 \sigma_{E_1} \mathbb{E}\big[\lambda\big(\gamma_1^T Z_1 / \sigma_{V_1}\big) \,|\, H_1 > 0\big] - \rho_0 \sigma_{E_0} \mathbb{E}\big[\lambda\big(\gamma_0^T Z_0 / \sigma_{V_0}\big) \,|\, H_0 > 0\big]. \tag{17}$$

This comprises the following four elements:

$$\rho_1 \sigma_{E_1} \int \big[\lambda\big(\gamma_1^T z / \sigma_{V_1}\big) \Phi_1\big(\gamma_1^T z / \sigma_{V_1}\big) - \lambda\big(\gamma_0^T z / \sigma_{V_0}\big) \Phi_1\big(\gamma_0^T z / \sigma_{V_0}\big)\big]\, dF_{Z_1}(z)$$

$$+ \rho_1 \sigma_{E_1} \bigg[\int \lambda\big(\gamma_0^T z / \sigma_{V_0}\big) \Phi_1\big(\gamma_0^T z / \sigma_{V_0}\big)\, dF_{Z_1}(z) - \int \lambda\big(\gamma_0^T z / \sigma_{V_0}\big) \Phi_0\big(\gamma_0^T z / \sigma_{V_0}\big)\, dF_{Z_0}(z)\bigg]$$

$$+ \rho_1 (\sigma_{E_1} - \sigma_{E_0}) \mathbb{E}\big[\lambda\big(\gamma_0^T Z_1 / \sigma_{V_0}\big) \,|\, H_0 > 0\big]$$

$$+ \sigma_{E_0} (\rho_1 - \rho_0) \mathbb{E}\big[\lambda\big(\gamma_0^T Z_1 / \sigma_{V_0}\big) \,|\, H_0 > 0\big], \tag{18}$$

where $\Phi_k(\gamma_r^T z / \sigma_{V_r}) = \Phi(\gamma_r^T z / \sigma_{V_r}) / \int \Phi(\gamma_r^T z / \sigma_{V_r})\, dF_{Z_k}(z)$ is the counterfactual probability of selection in group k when the selection rule is as in group r and Φ denotes the standard normal CDF. The first two elements in (18) capture the effect of changes over time in the observable characteristics of the selected population. The first results from applying the selection rule from period 1 to period 0 holding the composition of period 1 fixed. The second captures changes in the distribution of characteristics from period 1 to period 0. The third element captures the effect of changes over time in the composition of the selected population in terms of unobservables through the variance in wages. The fourth element is a sorting effect that captures changes over time in the composition of the selected population in terms of unobservables through the correlation coefficient.[9] In our view, the first element belongs to the selection effect while the second element belongs to the composition effect as it is driven by changes over time in the distribution of Z. It is also not clear why the change in the variance of wages should be interpreted as a selection effect as it captures factors potentially unrelated to selection. As the final element reflects changes how unobservables are valued, it is arguably a component of the selection effect. However, it could be assigned the interpretation of a structural effect as it captures the market value of unobserved characteristics. We agree with the MR interpretation that the sign of ρ_t captures the nature of sorting.

We now present the selection, composition, and structural effects for our decomposition. Plugging the expression for $m_t(x, v)$ into (16) gives, after some straightforward calculations:

$$m^s_{Y_{\langle t,k,r \rangle}} = \int \big[\alpha_t^T x + \rho_t \sigma_{E_t} \lambda\big(\gamma_r^T z / \sigma_{V_r}\big)\big] \Phi_k\big(\gamma_r^T z / \sigma_{V_r}\big)\, dF_{Z_k}(z).$$

---

[9]MR make the strong assumption that the distribution of the covariates does not change over time, that is, $F_{Z_0} = F_{Z_1}$, so the second component drops out.

Our selection effect is

$$m^s_{Y_{\langle 1,1,1 \rangle}} - m^s_{Y_{\langle 1,1,0 \rangle}}$$

$$= \alpha_1^T \int x \big[\Phi_1(\gamma_1^T z/\sigma_{V_1}) - \Phi_1(\gamma_0^T z/\sigma_{V_0}) \big] \, dF_{Z_1}(z)$$

$$+ \rho_1 \sigma_{E_1} \int \big[\lambda(\gamma_1^T z/\sigma_{V_1}) \Phi_1(\gamma_1^T z/\sigma_{V_1}) - \lambda(\gamma_0^T z/\sigma_{V_0}) \Phi_1(\gamma_0^T z/\sigma_{V_0}) \big] \, dF_{Z_1}(z). \quad (19)$$

The first element on the right-hand side of (19) is the effect on the average wage from changes in the distribution of observable characteristics of the selected population, holding the population distribution constant, resulting from applying the selection equation from period 0 to period 1. It is positive if those entering the selected population have characteristics associated with higher wages. This element is missing in the selection effect in (18). The second element is the corresponding effect for the unobservable characteristics and corresponds to the first in (18).

Our composition effect is

$$m^s_{Y_{\langle 1,1,0 \rangle}} - m^s_{Y_{\langle 1,0,0 \rangle}}$$

$$= \alpha_1^T \left[\int x \, \Phi_1(\gamma_0^T z/\sigma_{V_0}) \, dF_{Z_1}(z) - \int x \, \Phi_0(\gamma_0^T z/\sigma_{V_0}) \, dF_{Z_0}(z) \right]$$

$$+ \rho_1 \sigma_{E_1} \left[\int \lambda(\gamma_0^T z/\sigma_{V_0}) \Phi_1(\gamma_0^T z/\sigma_{V_0}) \, dF_{Z_1}(z) \right.$$

$$\left. - \int \lambda(\gamma_0^T z/\sigma_{V_0}) \Phi_0(\gamma_0^T z/\sigma_{V_0}) \, dF_{Z_0}(z) \right]. \quad (20)$$

The first element on the right-hand side of (20) is the change in the average wage resulting directly from changes over time in the distribution of the observable characteristics while the second element is the same as the second term in (18).

Finally, our structural effect is

$$m^s_{Y_{\langle 1,0,0 \rangle}} - m^s_{Y_{\langle 0,0,0 \rangle}}$$

$$= (\alpha_1 - \alpha_0)^T \mathbb{E}[X_0 | H_0 > 0] + (\rho_1 \sigma_{E_1} - \rho_0 \sigma_{E_0}) \mathbb{E}\big[\lambda(\gamma_0^T Z_0/\sigma_{V_0}) \, \big| \, H_0 > 0 \big]. \quad (21)$$

The first element on the right-hand side of (21) reflects the impact of changes over time in the returns to observable characteristics while the second captures the type and degree of sorting and is the same as the sum of the third and fourth elements in (18). As the expectation involving the inverse Mills ratio is positive, its contribution is positive when $\rho_1 \sigma_{E_1} > \rho_0 \sigma_{E_0}$.

Finally, we present a simple example illustrating that the two elements of the structural effect cannot generally be identified in nonseparable models.[10] Consider a multi-

[10]Chernozhukov, Fernández-Val, and Luo (2019) separate the structural effect from the sorting effect using an exclusion restriction in the copula for the unobservables of the employment and wage equation.

plicative version of the parametric HSM obtained by replacing (14) with

$$Y_t = \alpha_t^T X_t E_t, \quad \text{if } H_t > 0,$$

and weaken the parametric assumption on the joint distribution of E_t and V_t by only requiring that $V_t \sim N(0, 1)$ and $\mathbb{E}[E_t \mid Z_t, H_t > 0] = \rho_t \lambda(\gamma_t^T Z_t)$. α_t and ρ_t cannot be separately identified from the moment condition:

$$\mathbb{E}[Y_t \mid Z_t, H_t > 0] = \alpha_t^T X_t \rho_t \lambda(\gamma_t^T Z_t).$$

4. Empirical implementation

We briefly describe our empirical implementation to obtain the counterfactuals. A more detailed description is provided in the Online Appendix D. Although there are differences between the different models, each requires three steps: (1) estimation of the control function(s), (2) estimation of the LDSF for wages, and (3) estimation of the counterfactual distribution. All the estimators use flexible parametric specifications as in Chernozhukov, Fernández-Val, Newey, Stouli, and Vella (2020).

We start with the censored selection model in Section 3.1. Step (1) requires preliminary estimates of the probability of censoring. We assume a binary response model for the probability of censoring $\pi(z) = \Lambda(P(z)^T \pi_0)$, where $\Lambda(\cdot)$ is a link function and $P(\cdot)$ is a vector of transformations with good approximation properties. As we use the logistic function as the link function throughout, this implies that π_0 can be estimated via logit and $\widehat{\pi}(z) = \Lambda(P(z)^T \widehat{\pi}_0)$ is the plugin estimator of $\pi(z)$. We assume a distribution regression model for $F_{H^* \mid Z}$, (i.e., $F_{H^* \mid Z}(h \mid z) = \Lambda(R(z)^T \gamma(h))$), where $R(\cdot)$ is a vector of transformations with good approximation properties. This enables us to use a censored distribution regression to estimate $\gamma(h)$ via a series of logits. The estimator of the control function is given by $\widehat{V}_i = \Lambda(R(Z_i)^T \widehat{\gamma}(H_i))$, where $\widehat{\gamma}(H_i)$ is the estimator of $\gamma(h)$ at $h = H_i$. For the second step estimation of the LDSF for wages, we assume that $G(y, x, v) = \Lambda(Q(x, v)^T \beta(y))$, where $Q(x, v)$ is a vector of transformations of (x, v) with good approximating properties and we estimate $\beta(y)$ via distribution regression on the observed wages of the selected sample through a series of logits. The final step substitutes the parameters with their estimators to obtain the sample analog of (5):

$$\frac{1}{n^s} \sum_{i=1}^{n^s} \Lambda(\widehat{Q}_i^T \widehat{\beta}(y)) \, 1\{\widehat{V}_i > 1 - \widehat{\pi}(Z_i)\},$$

where n^s is the number of observations in the selected sample and $\widehat{Q}_i = Q(X_i, \widehat{V}_i)$.

Estimation of the model with the double selection mechanism is similar to the single censoring rule except that we require the estimation and use of two control functions.

They provide a wage decomposition with 4 components: structural, composition, selection, and sorting. Compared to our decomposition, the composition and selection effects are the same, but our structural effect corresponds to the sum of their structural and sorting effects.

We assume binary response models for the probabilities that H and W are censored:

$$\pi_H(z) := \mathbb{P}(H > 0 \,|\, Z = z) = \Lambda\big(P_H(z)^T \pi_{H,0}\big),$$

$$\pi_W(z) := \mathbb{P}(W > 0 \,|\, Z = z) = \Lambda\big(P_W(z)^T \pi_{W,0}\big),$$

where $P_H(\cdot)$ and $(P_W(\cdot))$ are vectors of transformations. The coefficient vectors $\pi_{H,0}$ and $\pi_{W,0}$ are estimated by logit. We again assume distribution regression models for $F_{H(Z,V_H)|Z}$ and $F_{W(Z,V_H)|Z}$. That is,

$$F_{H(Z,V_H)|Z}(h \,|\, z) = \Lambda\big(R_H(z)^T \gamma_H(h)\big),$$

$$F_{W(Z,V_W)|Z}(w \,|\, z) = \Lambda\big(R_W(z)^T \gamma_W(w)\big),$$

where $R_H(\cdot)$ and $R_W(\cdot)$ are vectors of transformations. The coefficient vectors $\gamma_H(h)$ and $\gamma_W(w)$ are estimated accordingly. The estimators of the control functions, LDSF and counterfactuals are similar to those for the censored regression model.

The estimation of the ordered selection model differs from the censored selection models. First, we assume

$$\mu_h(z) = \mathbb{P}(H \leq h \,|\, Z = z) = \Lambda\big(P(z)^T \gamma_h\big),$$

for $0 \leq h < K$ and where $P(\cdot)$ is a vector of transformations. We estimate γ_h using a variant of the ordered logit model, which allows the γ_h's to depend on h. The estimators of the control functions μ_h are calculated by plugging the estimated γ_h's into the equations above. Estimation of the LDSF relies on the following additional assumption:

$$\int_{\mu_{h-1}(z)}^{\mu_h(z)} G(y, x, v)\, dv = \Lambda\big(Q(x, \mu_h, \mu_{h-1})^T \beta(y)\big)\big(\mu_h(z) - \mu_{h-1}(z)\big),$$

for $h = 1, \ldots, K$ with $\mu_K(z) = 1$ and where $Q(\cdot, \cdot, \cdot)$ is a vector of transformations. The vector $\beta(y)$ can be estimated by distribution regression. The counterfactual distribution can be estimated by employing the sample analog of (13).

Following FVV, inference is based on the weighted bootstrap (Praestgaard and Wellner (1993)). This method obtains the bootstrap version of the estimator of interest by repeating all the estimation steps including sampling weights drawn from a nonnegative distribution with mean and variance equal to one (e.g., standard exponential).

5. Empirical results

5.1 *Hours equation*

We start by describing the variables included in Z using the ASEC. Following MR, we include six indicator variables for the highest educational attainment reported. Namely, (i) 0–8 years of completed schooling, (ii) high school dropouts, (iii) high school graduates or 12 years of schooling, (iv) some college, (v) college, and (vi) advanced degree. We include a quartic polynomial in potential experience and interact this with the education

levels.[11] We use 5 indicator variables for marital status: (1) married, (2) separated, (3) divorced, (4) widowed, and (5) never married. We include indicator variables for Black and Hispanic and four indicator variables for regions: northeast, midwest, south, and west. Finally, we use linear terms for the number of children aged less than 5 years interacted with the indicator variables for marital status. For the MORG, we use 5 levels of education as the two lower categories in the ASEC are merged. The variables Black, Hispanic, experience, and region are the same. Only one indicator for marital status is used (married or not) and we employ household size, and its interaction with marital status, as the only household characteristic.

With the exception of the household size and composition variables, all of the conditioning variables appear as both determinants of annual hours and hourly wages. While one might argue that household size and composition may affect hourly wage rates, we regard these exclusion restrictions as reasonable and note that similar restrictions have been previously employed (see, e.g., MR). However, given their potentially contentious use we explore the impact of not using them below. The assumption that annual hours of work do not affect the hourly wage rate means that the variation in hours across individuals is a source of identification.

Although our primary focus is the wage decomposition, we highlight the major features of the hours equation estimates. The hours equation is estimated by distribution regression and we report the marginal impact of being below a certain point in the annual hours distribution. We do this for education level, race, marital status, and for one of the exclusion restrictions (having a child below the age of 5). We only report the result of the latter here; see Figure 5. The selected points in the hours distribution are 0, 1000, and 2000. We find that many of the individual characteristics have an impact on the level of annual hours worked. This is not particularly surprising given the large literature on labor supply documenting the roles of education and marital status on labor market participation. Perhaps what is more surprising is that the magnitude of the impact of these variables does not appear to change substantially over the sample period in either the ASEC or MORG data. The exception is with respect to the exclusion restrictions which became less important over time. This is consistent with Card and Hyslop
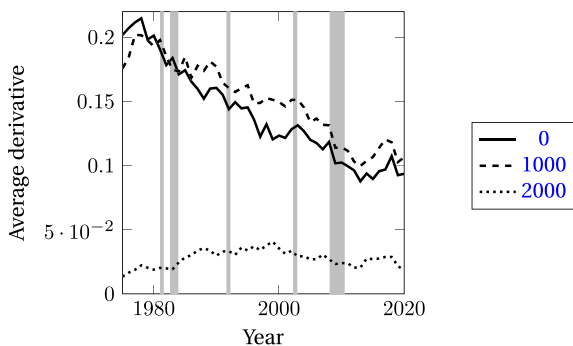

FIGURE 5. Average derivative of the exclusion restriction at different levels of hours.

[11]We employ the methodology described in MR for education and potential experience.

(2021). Note that the level of education has drastically increased over the sample period and this has had a substantial impact on the hours distribution.

We also estimated models for annual weeks of work using distribution regression and ordered models for annual weeks and annual hours using the ASEC. The flavor of the results are very similar to those for the annual hours equations. The individual variables which also appear in the wage equations are statistically significant and their impact does not appear to change a great deal over the sample period. The variables employed as exclusion restrictions have a statistically significant impact on the level of reported work.

5.2 *Decompositions*

The wage equations are estimated for each year by distribution regression over the subsample with positive hourly wages. The conditioning variables are those in the hours equation with the exception of the household size and composition variables. We also include the appropriate control function, its square, and interactions between the other conditioning variables and the control function. The decompositions require a base year. The steadily increasing female participation rate suggests that 1975 is a reasonable choice as it has the lowest level of participation. It is reasonable to assume that those with a certain combination of x and v working in 1975 also have a positive probability of working in other years as required by the identification conditions. We present results for the censoring mechanism being either the number of annual hours or the number of annual weeks for the ASEC. These are shown in Figures 6 and 7, respectively.[12] The results for the MORG, using numbers of hours in reference week as the censoring mechanism, are in Figure 8. Before presenting the decompositions we highlight that they reflect the change in the contributions of each component over time. For example, if the structural effects are zero this should be interpreted as the contribution of the structural effects not changing, not that there are no structural effects.

We start with annual hours as the censoring mechanism and Figure 6A presents the decomposition for the mean, which increases by 25% over our sample period. The total effect is driven by the composition effect although in several instances the structural effect is contributing. It is generally negative and small relative to the composition effect. The contribution of the selection component is negative and small. A negative selection component implies that females are positively selected into employment and those who entered employment between 1975 and 2020 were less productive than those already employed.

Figures 6B–6F present the decompositions for the 10th, 25th, 50th, 75th, and 90th percentiles. A number of important conclusions can be drawn. The large increases at each quantile are driven by the large changes in the composition effects. This reflects the increasing education levels of the female workforce. The large increases from the composition effects at each quantile are somewhat offset by the structural effects at lower

---

[12]Note that the line depicted as "Total" in the figures represents the total change in wages measured as the sum of the various estimated components. Estimation error may make this vary from the actual change in wages.
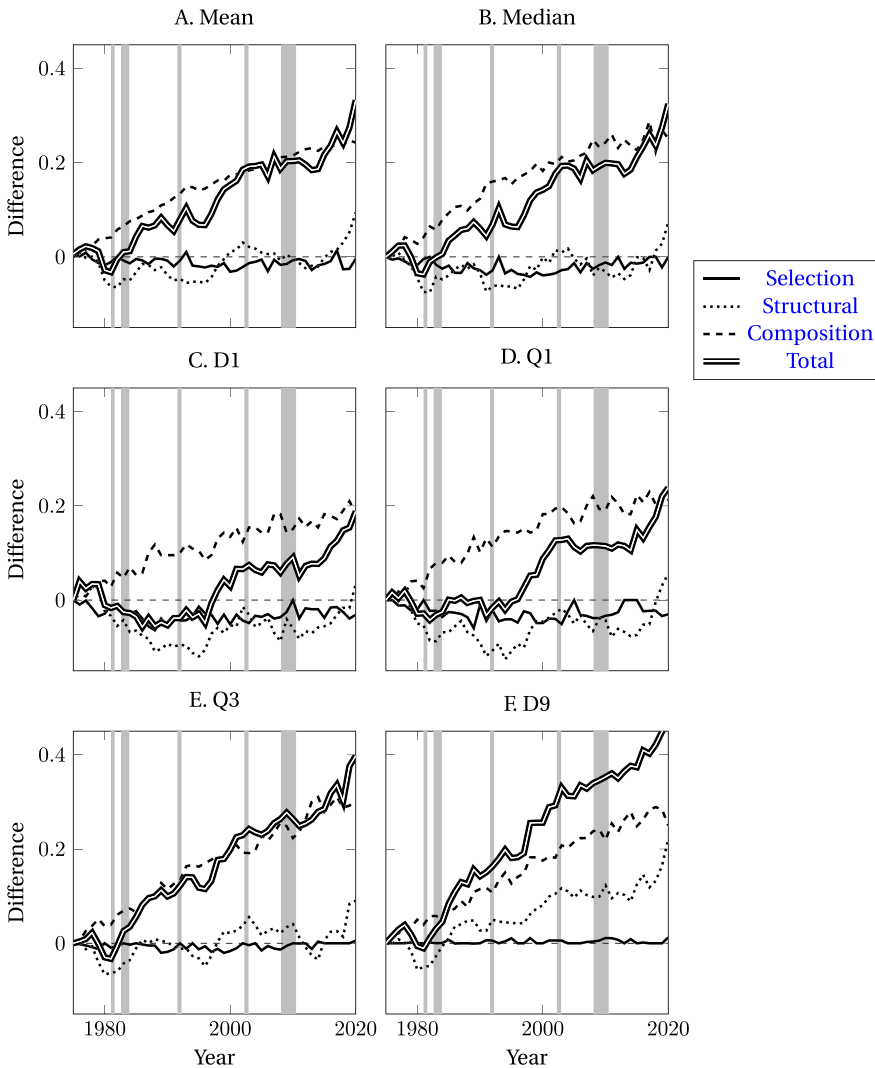
FIGURE 6. Decompositions using the ASEC data and annual working hours.

quantiles and there is only clear evidence of an economically important and positive structural effect at D9. At this quantile almost half of the 44% increase in real wages experienced over the sample period is due to the structural effect. At lower quantiles, the structural effects are somewhat cyclical while the sustained increase at D9 reflects the increasing skill premia.

Figures 6B–6F reveal no indication of changes in the selection effects at Q3 and D9. Individuals located here are likely to have had a relatively strong commitment to employment in 1975 and there have been no substantial movements in their hours distribution. Moreover, these individuals are less likely to incur periods of unemployment. Lower down the wage distribution there is evidence of small negative changes in selection during some years in our sample. At D1 and Q1, the selection effects are neg-
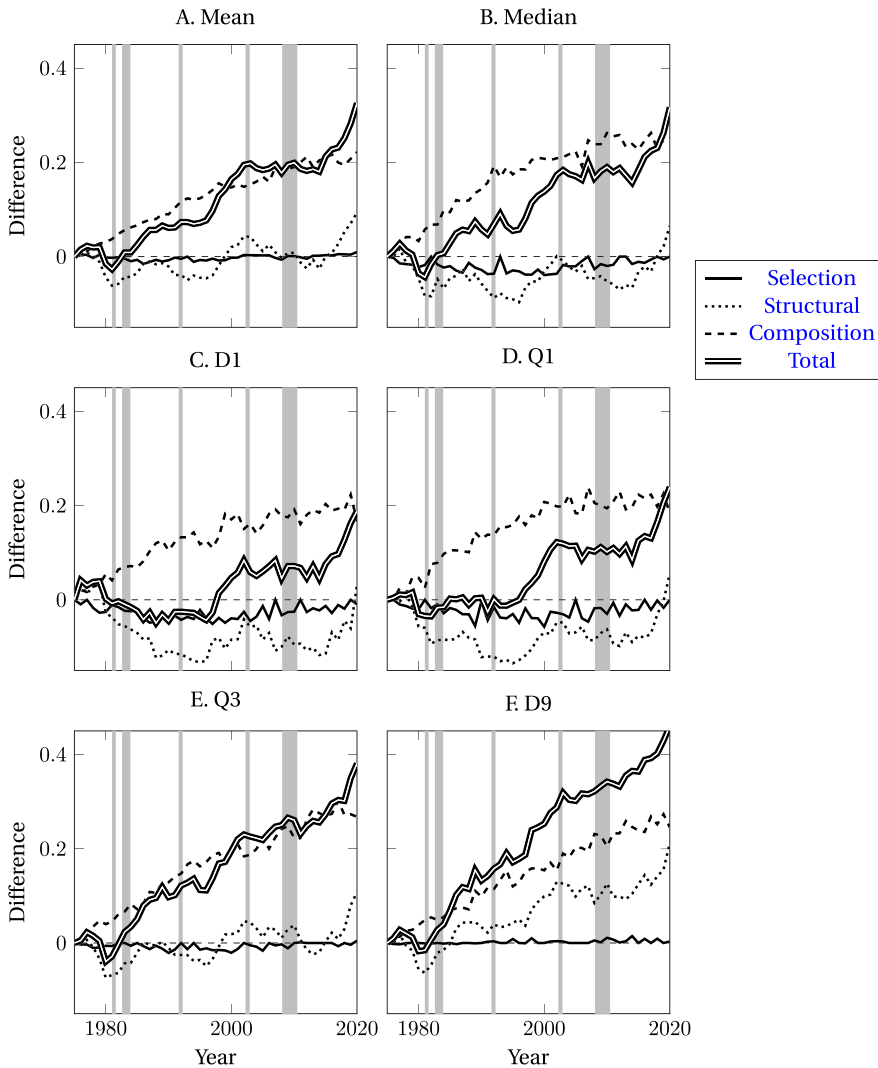
FIGURE 7. Decompositions using the ASEC data and annual working weeks.

ative and economically important. The confidence intervals for these selection effects are presented in Figure 9. They indicate that for several time periods selection effects are statistically significantly different from zero. At D1, the total real wage increase is less than 10% for the majority of the sample period and the negative selection effect is frequently of the order of around 3%.

We explore the form of sorting implied by these results. Given the nonseparable nature of the model, this is not straightforward although one indication of the sorting pattern is how wages change at each quantile in response to a change in the control function. This corresponds to the local average derivative function of FVV evaluated at each quantile. This is somewhat comparable to the coefficient on the selection term in the HSM. A positive average derivative at a specific quantile suggests that the unobserv-
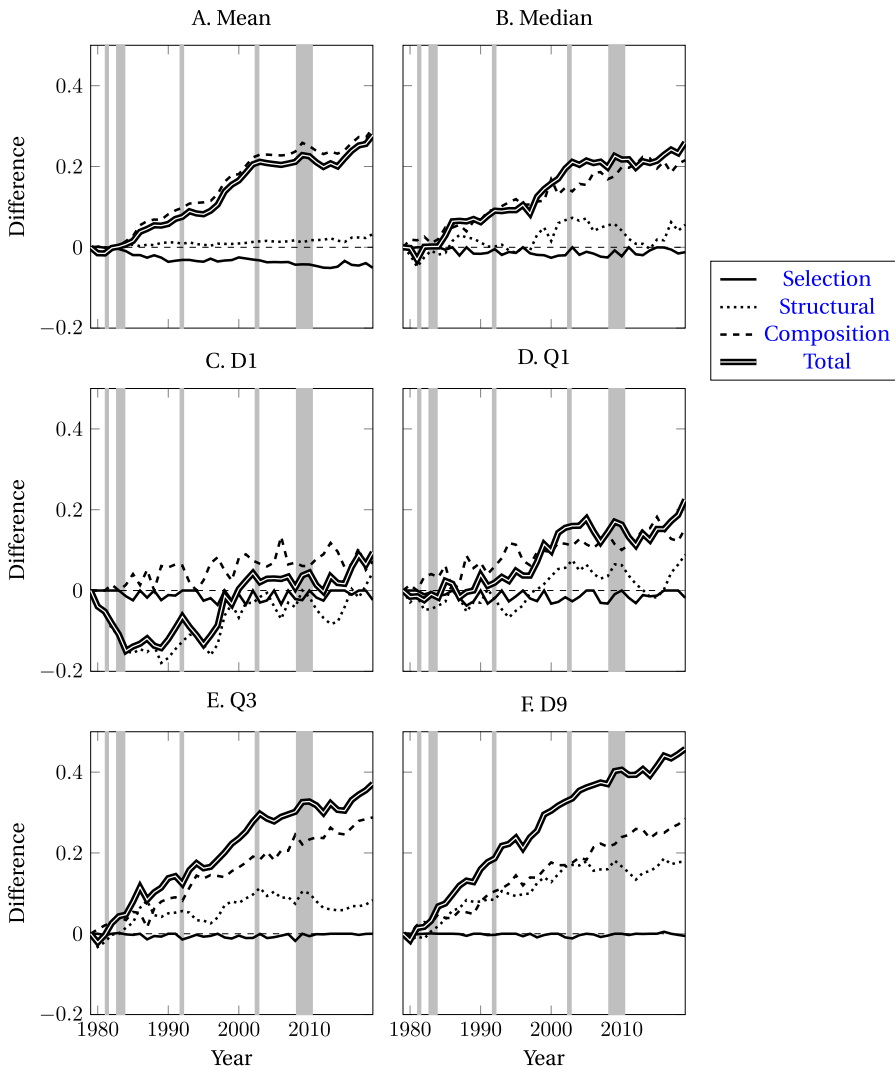
FIGURE 8. Decompositions using the MORG data and annual working hours.

ables increasing hours are positively correlated with wages at that quantile. The results are shown in Figure 10. At D1, Q1, and Q2, these estimated derivatives are positive and large for the whole sample period. At Q3, the derivative is positive for the vast majority of the sample period although there are instances where it is close to zero or very slightly negative. At D9, it is negative but small in magnitude. The results clearly support the existence of a positive relationship between V and wages at the bottom of the distribution implying positive selection, which is more important in the lower half of the wage distribution. This is similar to the findings of AB for female wages in the British labor market for 1978 to 2000.

Annual hours is an economically attractive censoring mechanism as it exploits the variation in annual hours induced both by hours and by weeks. However, it is possible
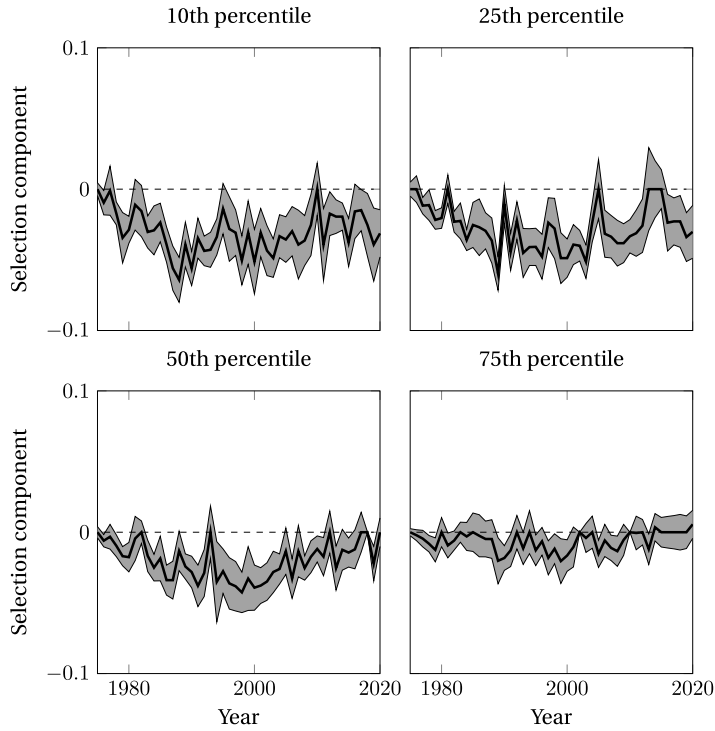
FIGURE 9. Selection components and associated 95% confidence intervals for females at various percentiles.

that selection operates either through hours or weeks exclusively. We first address this issue by replacing annual hours with annual weeks as the selection mechanism. The results from these decompositions using the ASEC are in Figures 7A–7F. Their primary feature is their similarity to those for annual hours. This suggests that the control function from the annual hours censoring mechanism is highly correlated with that from annual weeks despite the differences in their respective distributions.
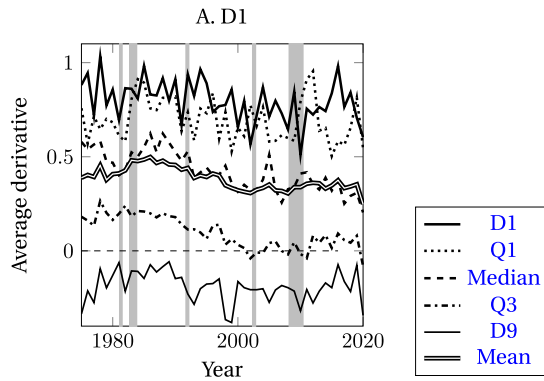

FIGURE 10. Average derivatives at different quantiles.

Now consider the decompositions for the MORG recalling that wages are measured differently than in the ASEC and the hours measure is based on the survey week. We implement our censored selection estimator using hours in the survey week as the censoring mechanism noting that only a subset of the exclusion restrictions used in the ASEC are available in the MORG. The results are shown in Figures 8A–8F. As the MORG covers a different sample period, the figures look slightly different from those for the ASEC. Nevertheless, the findings regarding the role of selection are almost identical. The similarity across figures is remarkable given the differences in measurement, data, and exclusion restrictions.

We acknowledge that although also employed by MR and MW, the use of household composition variables as exclusion restrictions is controversial. Accordingly, we reproduced the decompositions from the censored selection mechanisms first excluding the household variables from both the hours and wage equations and then including them in both equations. The model is now only identified by the variation in the number of hours worked. We do not find any remarkable changes from either model, in comparison to the specification employed above, with respect to the presence or magnitude of selection effects. The only notable difference is the presence of occasionally larger negative selection effects at the bottom decile for the specification, which excludes the family composition variables from both equations.

5.3 *Results of the double selection model*

Our results from Section 5.2 seem robust to the use of either hours or weeks as the selection variable in the censored selection model. Figures 11A–11C report the decomposition for the double selection mechanism. There continues to be no evidence of selection above the median so we report the decompositions for D1, Q1, and the median.

While there are some differences in these figures compared to those for selection using only annual hours or annual weeks they are relatively small. These results seem to suggest that the unobservables, which increase participation on any margin, such as usual weeks, usual hours, hours in survey week, are all highly correlated. To pursue this possibility, we estimate the average derivative with respect to each of the control functions evaluated at the mean wage as this provides some insight into the source of selection. The derivatives, reported in Figure Web-1 in the Online Appendix, indicate that at the mean wage both sources of selection are important. The derivative for the weeks' control function is the bigger of the two and the effect is relatively constant with the exception of a notable decrease at the time of the financial crisis. The hours' control function derivative is negative for the earliest years of the sample before turning and remaining positive for the remaining years. Prior to the financial crisis it increases in magnitude and for a very short period it is the larger of the two. The high correlation between the two control functions makes it difficult to interpret this figure but this evidence suggests that selection operates both through the weeks and hours decisions and the effect of each is similar.
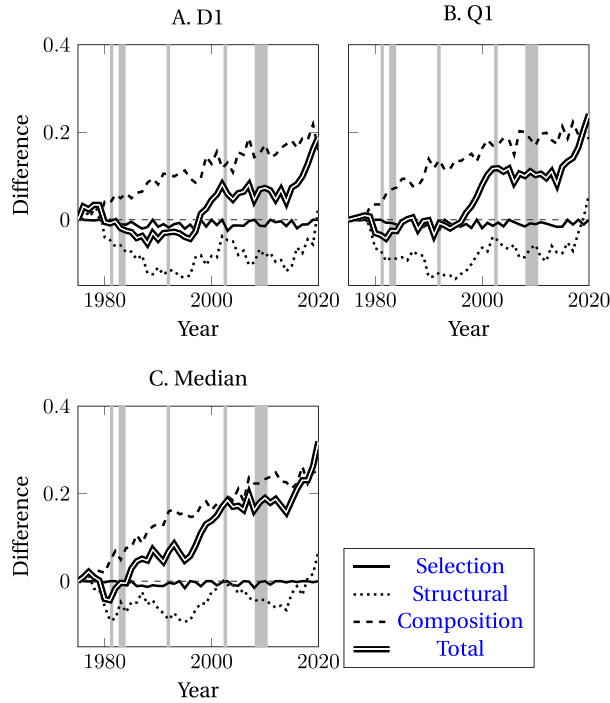
FIGURE 11. Decompositions using a double selection mechanism based on ASEC data.

5.4 *Results of the ordered selection model*

Figure 12 reports the decomposition exercise using the model in Section 3.4 incorporating bunching in the selection variable. We first employ the ASEC with annual weeks as the ordered selection rule. The separation values are 0, 10, 20, 30, 40, and 50 weeks. The decompositions are very similar to those for the continuous selection rule. The most notable feature of the decompositions are the selection effects at D1 and Q3. These effects are slightly larger than those at the corresponding quantiles in the earlier figures and represent important economic effects. There is clear evidence that the changes in selection effects have decreased wages at this quantile in some time periods. We also employed annual hours as the ordered selection variable using 0, 500, 1000, 1500, 2000, and 2500 annual hours as the separation values. These results, available from the authors upon request, are very similar to those using ordered weeks. The selection effects are very similar to those in Figure 12 although the allocation of the changes across the structural and composition effects differs somewhat.

6. COMPARISON WITH MULLIGAN AND RUBINSTEIN (2008)

MR examined changes in the mean female full-time wage for white females for the period 1975–1999 and found negative sorting effects in the 1970s. These sorting effects became positive in the early 1980s and by the late 1990s they were economically large. The change from negative to positive sorting is inconsistent with our findings and we at-
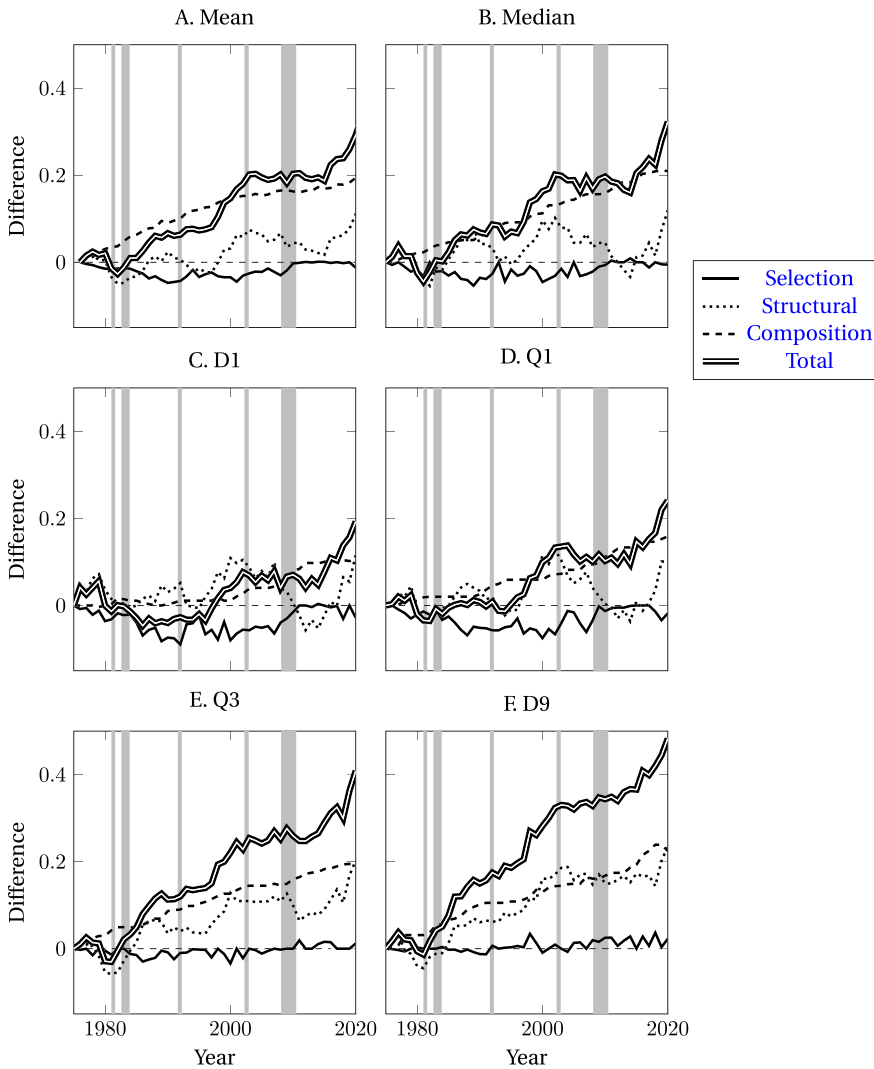
FIGURE 12. Decompositions using the ASEC data and annual working weeks using our alternative model taking account of bunching.

tempt to reconcile these contrasting conclusions. Some of the discussion which follows refers to figures provided in the Online Appendix E.

We first explore the behavior of the term $\rho_t \sigma_{E_t}$ in (17). We estimate the HSM using the MR sample and exclusion restrictions to obtain the results shown in Figure Web-2A. To ensure the differences across our and the MR results are not driven by their focus on FTFY white females, we repeat the exercise including nonwhites and all those reporting positive working hours. This is reported in Figure Web-2B. These figures confirm the MR results. For both samples, $\rho_t \sigma_{E_t}$ is negative at the beginning of the sample but becomes positive and economically large during the sample period. The results imply that the changes in the selection effects, as defined by MR, are large.

Figure Web-3 presents the selection effect based on our sample using the estimates from the HSM. The figure decomposes the total selection effect into the four components shown in (18), which reflect the components in the decomposition from the HSM. It reveals that the changes in the selection effects are almost entirely due to the change in the estimate of ρ_t. The impact of the unobservables from the selection equation on wages was initially negative but became positive and subsequently large in magnitude. Figure Web-4 presents the selection, structural, and composition effects from our estimation approach based on (19)–(21). We find that our selection effect is small and the majority of the wage change reflects composition effects. Figure Web-4 is similar to Figure 6A obtained via our approach. The reason for our relatively smaller selection effect is clear from Figure Web-3. The figure reveals that the second component, which from (18) is included in our selection effect, is small and similar in size to that of Figure 6A and Figure Web-4. This suggests that the differences in results from our approach and MR's use of the HSM is partially due to the difference in definitions recalling this reflects the inability to isolate the MR selection effects in nonseparable models.

Given that the return to observable skills has increased over the sample period, it is not controversial to find that the estimate of ρ_t has also become large. However, the result that ρ_t changes sign has implications for the nature of sorting into the labor market. As ρ_t captures the mapping from unobservables in the work equation to wages, we examine whether our approach uncovers the same pattern. We do so by evaluating again the average derivative for wages with respect to the control function as presented in Figure Web-5A. This is a measure of how unobservables driving the hours decision affect wages. While the derivative's value at the mean fluctuates over the sample period, it does not change sign and it is always consistent with positive sorting.

As noted above, this positive sorting result over the whole sample period contrasts with the conclusions of MR. To examine why this difference might occur, we consider differences between the HSM and our methodology which may be responsible. The three obvious possible causes are: (i) the additional identifying power introduced through hours as a censoring variable in the selection equation in our approach in contrast to the MR reliance on exclusion restrictions and distributional assumptions; (ii) the use of the normality assumption in the HSM, which is not employed in our approach; and (iii) the nonseparable nature of our model.

We address each of these issues in turn. We first examine the impact of the additional source of identification by estimating the model using a parametric approach, which relies on normality and separability but which exploits the variation in hours for identification purposes. We employ the Vella (1993) procedure, which estimates the hours equation parameters by Tobit and computes the generalized residual, defined for the $H > 0$ observations as $H - Z^T \widehat{\gamma}$ where $\widehat{\gamma}$ is estimated by Tobit, to include as the correction for selection in the wage equation. This assumes that hours do not directly effect hourly wage rates.[13] To make the estimates more comparable to those from the HSM, we divide the generalized residual by the estimated standard deviation of working hours, σ_{V_t}. The only difference between this approach and the HSM is the use of the Tobit generalized

---

[13]Hirsch (2005) provides empirical evidence supporting this assumption.

residual rather than the inverse Mills ratio as the correction for selection. We plot the corresponding coefficient on the Tobit generalized residual in Figure Web-5B recalling that this coefficient estimates $\rho_t \sigma_{E_t}$.

Two striking features are revealed from an examination of Figure Web-5B. First, recall that under normality the estimates of $\rho_t \sigma_{E_t}$ and the coefficient on the generalized residual should be identical. However, the estimates are very different. More importantly, the coefficient estimate for the Tobit generalized residual is always positive. As there is no reason that departures from normality will bias the estimates of $\rho_t \sigma_{E_t}$ and the coefficient for the Tobit generalized residual in the same manner, one could interpret the difference in the estimates as evidence of nonnormality. However, as the Tobit generalized residual also exploits variation in the hours variable for identification purposes it is also possible that this is the source of, or at least contributes to, the difference in the signs and the behavior of the two coefficients. Second, it also worth noting that the pattern of movement in the coefficient on the Tobit generalized residual is almost identical to the average derivative of our control function despite the drastically different ways in which each is computed. Moreover, while the use of the Tobit generalized residuals relies on separability, the estimation of the average derivative does not. Thus the difference between the HSM results and ours are not due to the relaxation of separability. The two procedures, namely our procedure and that using the Tobit generalized residual, are very different but each exploits the variation in hours as a means of identification. This is strongly suggestive that the use of hours as an additional source of identification is driving the difference in the implications for sorting between the HSM and our approach.

While it appears that the use of the variation in hours as the source of identification is the cause of the differences with MR, it is possible that the departures from normality may also be responsible. The final approach we explore is the use of the propensity score as the control function noting that we allow it to enter the wage equation in a nonseparable manner (see, e.g., Newey (2007)). The propensity score employs the exclusion restrictions as the sole source of identification. We estimated the model and computed the average derivative of wages with respect to the propensity score. The results are presented in Figure Web-5C. This derivative also changes sign as we move through the sample period and shows behavior similar to the estimate of ρ_t from the HSM. We conclude that the differences in terms of the relationship between E_t and V_t between our results and MR are due to the use of the variation in hours as a source of identification.

For reconciling the difference in the results associated with the two generalized residuals in the separable setting and assuming normality, it is useful to further examine their sources of variation. Recall that for the subsample of workers, the variation in their values of the inverse Mills ratio is due to the variation in Z. In contrast, an individual's value of the Tobit generalized residual also exploits the value of H. Now consider the case where ρ_t is positive and there are two working individuals with identical Z but where one of the individuals receives a much larger positive value of V_t. This produces a relatively larger positive value of E_t resulting in that individual having relatively higher hours and wages. In this setting, the value of the inverse Mills ratio for both individuals will be the same while the Tobit generalized residual of the individual with the higher

value of H will be greater than the other individual. This suggests that the Tobit generalized residual is capturing information regarding "sorting" into hours, which is ignored by the inverse Mills ratio. Moreover, the inverse Mills ratio, unlike the Tobit generalized residual, is unable to explain the variation in wages across these two individuals.

While it appears that the above explains the different results associated with the two different approaches, it is important to explore why ρ_t might change sign over time for the models identified solely by exclusion restrictions via the HSM as reported by MR. Recall that a negative ρ_t implies that the working individuals with the lowest probabilities of participation should have relatively low observed wages among individuals with the same observed characteristics X, which determine wages. The reverse is true for a positive ρ_t. We explore the impact of the variables employed as exclusion restriction in Z in settings where their exclusion is contentious and when their impact on the participation decision varies over the time period examined. We do this by estimating a wage regression identical to the second stage of MR while replacing the inverse Mills ratio by one of the variables, which is employed as an exclusion. Namely a dummy variable for a child below the age of 5 years. The impact on wages of having a "young child" was negative until 1982 at which time it turned, and remained positive. This corresponded to a period, also reported by Card and Hyslop (2021), in which the magnitude of the negative impact of a "young child" on the employment decision decreased. While we acknowledge the presence of other ongoing factors, the change in the impact of "young child" on both wages and participation could generate a change in the estimated sign of ρ_t. For example, in the absence of other influences, the large positive influence of "young child" on the value of the inverse Mills ratio combined with negative correlation between "young child" and wages would produce a negative value of ρ_t. In contrast a decreasing effect of "young child" on participation would produce a relatively smaller value for the inverse Mills ratio and that, combined with the positive correlation between "young child" and wages, would produce a positive ρ_t.

We highlight that we consider the above discussion as suggestive rather than conclusive. Our objective is to consider the possible causes of the differences in the results from the use of the two control functions. The evidence suggests that part of the difference is due to the use of variation in hours as a source of identification for ρ_t. However, it is clear that the effect of the exclusion variables on the hours and work decisions is also an important factor in identifying ρ_t, and this has changed over time. Related to this last issue is the validity of the exclusion restrictions employed and how this validity has changed over the sample period. While the use of either of the control functions should produce consistent estimates of the sorting parameters when the model is correctly specified, the impact under misspecification is unclear.

7. Wage inequality

We noted above that despite the large literatures on the impact of selection on females' wages and the increase in female wage inequality, there are few papers which focus on both issues. The important exceptions are AB, MW, Chernozhukov, Fernández-Val, and Luo (2019), and Blau, Kahn, Boboshko, and Comey (2021), which evaluate the impact of selection by contrasting changes in the observed levels of wage inequality with
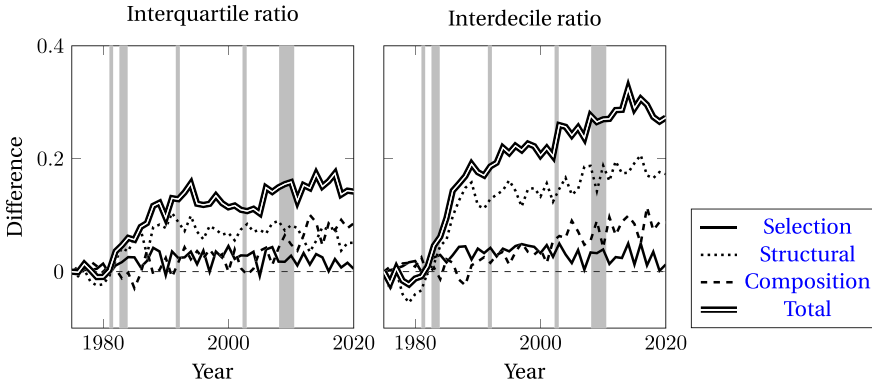
FIGURE 13. Decompositions of the interquartile and interdecile ratio using the ASEC.

the counterfactual levels associated with the total population working. The latter corresponds to a participation rate of 100%. We investigate the impact of selection by evaluating changes in the distributions, and also inequality, by holding the selection rule constant across time. While both approaches have merit, we prefer ours as participation rates do not approach 100% in the sample period.[14]

We provide the decompositions of changes in inequality using the annual hours as the censored selection variable for the ASEC data, hours in the survey week for the MORG and annual weeks as the ordered selection variable for the ASEC. For each of these models and selection rules, we decompose the interquartile and interdecile ratios. Those for annual hours using the ASEC are reported in Figure 13 and those for the MORG in Figure 14. The interquartile ratio is driven by each of the components. Neither the composition or structural effect dominates throughout the sample period. The selection effect contributes throughout the period and clearly increases inequality. The interdecile ratio is driven primarily by the structural effect especially during the drastic increase at the beginning of the sample period. The selection effect is clearly important
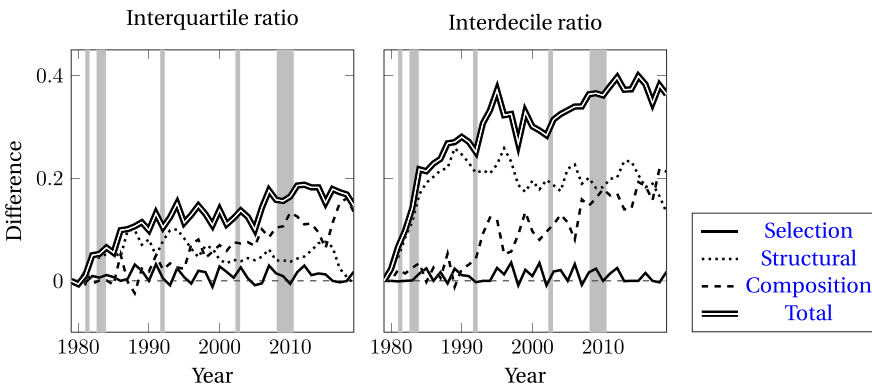

FIGURE 14. Decompositions of the interquartile and interdecile ratio using the MORG.

[14]Chernozhukov, Fernández-Val, and Luo (2019) considered both approaches.

and frequently more important than the composition effect. For the MORG, the conclusions regarding the structural and composition effects are similar to those for ASEC while the selection effects are slightly smaller. This reflects the smaller selection effects at lower quantiles in the wage decompositions (as presented in Figures 6 and 8). The evidence for both data sets support that selection has a modest but important impact on wage inequality that varies in magnitude over the sample period. As the wage decompositions based on the ordered selection rule suggested selection was more important than in the censored selection models (see Figure 12), we examine now whether this carries over to the inequality decompositions. We do not report the result but note that the evidence is similar to that for the censored selection rule.

Our results indicate that as an increasing number of females have entered the labor market they have reduced wages at the lower parts of the wage distribution while having no impact on wages above Q2. This increases measures of inequality based on ratios involving lower and upper quantiles. Potentially, there are two reasons why selection increases inequality based on whether either the observed or unobserved characteristics of those participating has changed over time. However, an examination of education levels, for example, suggests that observed characteristics have played a minor role. In particular, we find that those with education higher than high school degrees were more likely to participate over the whole sample period and that this did not change over time. This suggests that our results reflect changing unobserved characteristics. The selection effect captures the difference between the observed wage distribution and the counterfactual in which women participated as in 1975. Our decomposition method presented in (5) imposes that this difference captures the exit of females with lower levels of the control function. Figure Web-5, reveals a strong and positive relationship between wages and the control function suggesting that selection effects reflect that women entering the labor market were less productive than observationally identical women participating in 1975.

Our evidence of positive sample selection over the whole period implies that the decision to work is largely based on economic motivations. However, as employment rates have increased this has seen a reduction in sorting on economic grounds. This is consistent with the explanation provided by AB for the U.K. labor market. This is also consistent with the results above that the conventional household background family characteristics have become less important in explaining participation and hours worked. The reduction in positive sorting describes the changes in the hours distribution from the mid 1980s to the end of the 1990s. Blau, Kahn, Boboshko, and Comey (2021) argue that the booming economy and welfare reform may have played an important role in the 1990s. Our collective evidence suggests that post 2000 there was an increase in positive selection. This supports the evidence in Blau, Kahn, Boboshko, and Comey (2021). Toward the end of the sample period it appears that the impact of selection on inequality and, more generally, wages has returned to 1975 levels.

8. Conclusions

This paper documents the changes in female real wages over the period 1975 to 2020. We decompose these changes into structural, composition, and selection components

by estimating a nonseparable model with selection. 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 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 for the period 1975 to 2000 and this reflects a reduction in the level of positive sorting. However, post 2000 there appears to be an increase in positive sorting and the selection effects on wages and inequality appear to return to their 1975 levels by the end of our sample.

References

Acemoglu, Daron and David H. Autor (2011), "Skills, tasks and technology: Implications for employment and earnings." In *Handbook of Labor Economics*, Vol. 4B (David Card and Orley Ashenfelter, eds.), 1043–1171, Elsevier Science, North-Holland, Amsterdam. [572]

Angrist, Joshua, Victor Chernozhukov, and Ivan Fernàndez-Val (2006), "Quantile regression under misspecification, with an application to the U.S. wage structure." *Econometrica*, 74, 539–563. [572]

Arellano, Manuel and Stephanne Bonhomme (2017), "Quantile selection models with an application to understanding changes in wage inequality." *Econometrica*, 85, 1–28. [573]

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2008), "Trends in U.S. wage inequality: Revising the revisionists." *Review of Economics and Statistics*, 90, 300–323. [571, 572, 575]

Autor, David H., Allan Manning, and Christoffer L. Smith (2016), "The contribution of the minimum wage to US wage inequality over three decades: A reassessment." *American Economic Journal: Applied Economics*, 8, 58–99. [572]

Beaudry, Paul, David Green, and Benjamin M. Sand (2016), "The great reversal in the demand for skill and cognitive tasks." *Journal of Labor Economics*, 34, S199–S247. [577]

Blau, Francine D., Lawrence M. Kahn, Nikolai Boboshko, and Matthew L. Comey (2021), "The impact of selection into the labor force on the gender wage gap." Working Paper, Cornell University. [573, 601, 603]

Buchinsky, Moshe and Jinyong Hahn (1995), "Quantile regression model with unknown censoring point." Working Paper, Yale University, New Haven. [579]

Buchinsky, Moshe and Jinyong Hahn (1998), "An alternative estimator for the censored quantile regression model." *Econometrica*, 66, 653–671. [579]

Burdett, Kenneth and Dale T. Mortensen (1998), "Wage differentials, employer size and unemployment." *International Economic Review*, 39, 257–273. [580]

Card, David and Dean R. Hyslop (2021), "Female earnings inequality: The changing role of family characteristics on the extensive and intensive margins." *emphJournal of Labor Economics*, 39, S59–S106. [590, 591, 601]

Chernozhukov, Victor, Ivan Fernández-Val, and Siyi Luo (2019), "Distribution regression with sample selection, with an application to wage decompositions in the UK." [587, 601, 602]

Chernozhukov, Victor, Ivan Fernández-Val, and Blaisse Melly (2013), "Inference on counterfactual distributions." *Econometrica*, 81, 2205–2268. [572, 581]

Chernozhukov, Victor, Ivan Fernández-Val, Whitney K. Newey, Sami Stouli, and Frank Vella (2020), "Semiparametric estimation of structural functions in nonseparable triangular models." *Quantitative Economics*, 11, 503–533. [588]

Chernozhukov, Victor and Han Hong (2002), "Three-step censored quantile regression and extramarital affairs." *Journal of the American Statistical Association*, 97, 872–882. [579]

Cogan, John F. (1981), "Fixed costs and labor supply." *Econometrica*, 49, 945–963. [573, 579]

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux (1996), "Labor market institutions and the distribution of wages, 1973–1992: A semiparametric approach." *Econometrica*, 64, 1001–1044. [572, 580]

Fernández-Val, Iván, Aico van Vuuren, Francis Vella, and Franco Peracchi (2023), "Supplement to 'Selection and the distribution of female real hourly wages in the United States'." *Quantitative Economics Supplemental Material*, 14, https://doi.org/10.3982/QE1777. [574]

Fernández-Val, Ivan, Aico van Vuuren, and Frank Vella (2021), "Nonseparable sample selection models with censored selection rules." *Journal of Econometrics* (forthcoming). [573]

Fernández-Val, Ivan, Aico van Vuuren, Frank Vella, and Franco Peracchi (2022), "Hours worked and the U.S. distribution of real annual earnings 1976–2019." *Journal of Applied Econometrics* (forthcoming). [572]

Flood, Sarah, Miriam King, Steven Ruggles, and J. Robert Warren (2015), "Integrated public use microdata series, Current Population Survey: Version 4.0 [Machine-readable database]." Working Paper, University of Minnesota. [574]

Fortin, Nicole M., ThomasLemieux, and Sergio Firpo (2011), "Decomposition methods in economics." In *Handbook of Labor Economics*, Vol. 4A (David Card and Orley Ashenfelter, eds.), 1–102, Elsevier Science, North-Holland, Amsterdam. [581]

Heckman, James J. (1974), "Shadow prices, market wages and labor supply." *Econometrica*, 42, 679–694. [572]

Heckman, James J. (1979), "Sample selection bias as a specification error." *Econometrica*, 47, 153–161. [572, 573]

Heckman, James J., Jora Stixrud, and Sergio Urzua (2006), "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor Economics*, 24, 411–482. [572]

Heckman, James J. and Edward Vytlacil (2007), "Econometric evaluation of social programs, part II." In *Handbook of Econometrics, 6A* (James J. Heckman and Edward E. Leamer, eds.), 4878–5143, North-Holland. [573, 583, 584]

Hirsch, Barry T. (2005), "Why do part-time workers earn less? The role of worker and job skills." *Industrial and Labor Relations Review*, 58, 525–551. [599]

Imbens, Guido W. and Whitney Newey (2009), "Identification and estimation of triangular simultaneous equations models without additivity." *Econometrica*, 77, 1481–1512. [579]

Katz, Lawrence F. and David H. Autor (1999), "Changes in the wage structure and earnings inequality." In *Handbook of Labor Economics*, Vol. 3A (Orley Ashenfelter and David Card, eds.), 1463–1555, Elsevier Science, North-Holland, Amsterdam. [571]

Katz, Lawrence F. and Kenneth Murphy (1992), "Changes in relative wages, 1963–1987: Supply and demand factors." *Quarterly Journal of Economics*, 107, 35–78. [571]

Killingsworth, Mark R. (1983), *Labor Supply*. Cambridge University Press. [580]

Krueger, Allan and Eric Posner (2018), "A proposal for protecting low-income workers from monopsony and corruption." In *Revitalizing Wage Growth. Policies to Get American Workers a Raise* (Jay Shambaugh and Ryan Nunn, eds.), 139–156, Brookings, Washington, DC. [572]

Lee, David S. (1999), "Wage inequality in the United States during the 1980s: Rising dispersion or falling minimum wage?" *Quarterly Journal of Economics*, 114, 977–1023. [571, 572]

Lemieux, Thomas (2006), "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?" *American Economic Review*, 96, 461–498. [574, 575, 576, 577]

Maasoumi, Esfandiar and Le Wang (2019), "The gender gap between earnings distributions." *Journal of Political Economy*, 127, 2438–2504. [573]

Mulligan, Casey and Youna Rubinstein (2008), "Selection, investment, and women's relative wages over time." *Quarterly Journal of Economics*, 123, 1061–1110. [572, 597]

Murphy, Kenneth M. and Robert Topel (2016), "Human capital investment, inequality, and economic growth." *Journal of Labor Economics*, 34, S99–S127. [572]

Newey, Whitney K. (2007), "Nonparametric continuous/discrete choice models." *International Economic Review*, 48, 1429–1439. [583, 600]

Ñopo, Hugo (2008), "Matching as a tool to decompose wage gaps." *Review of Economics and Statistics*, 90, 290–299. [580]

Olivetti, Claudia and Barbara Petrongolo (2008), "Unequal pay or unequal employment? A cross country analysis of gender gaps." *Journal of Labor Economics*, 26, 621–654. [573]

Praestgaard, Jens and Jon A. Wellner (1993), "Exchangeably weighted bootstraps of the general empirical process." *Annals of Probability*, 21, 2053–2086. [589]

Vella, Frank (1993), "A simple estimator for models with censored endogenous regressors." *International Economic Review*, 34, 441–457. [599]

Welch, Finis (2000), "Growth in women's relative wages and inequality among men: One phenomenon or two?" *American Economic Review Papers & Proceedings*, 90, 444–449. [572]