

## Web Appendix of: Selection and the Distribution of Female Real Hourly Wages in the U.S.

IVÁN FERNÁNDEZ-VAL  
Boston University

AICO VAN VUUREN  
Gothenburg University, University of Groningen

FRANCIS VELLA  
Georgetown University

FRANCO PERACCHI  
University of Rome Tor Vergata

We analyze the role of selection bias in generating the 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 the selection bias from the employment decision by modelling 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 have increased wages at all quantiles while the impact of the structural effects varies by time period and quantile. Changes in the role of selection only appear at the lower quantiles of the wage distribution. The evidence suggests that there is positive selection in the 1970s which diminishes until the later 1990s. This reduces wages at lower quantiles and increases wage inequality. Post 2000 there appears to be an increase in positive sorting which reduces the selection effects on wage inequality.

KEYWORDS. wage inequality, wage decompositions, selection bias.

JEL CLASSIFICATION. C14, I24, J00.

### Web Appendix A: Control function with unknown censoring point

We provide an estimator of  $F_{H^*|X,Z}$  based on distribution regression. This is an alternative to Buchinsky and Hahn (1998) and Chernozhukov and Hong (2002), which

---

Iván Fernández-Val: [ivanf@bu.edu](mailto:ivanf@bu.edu)

Aico van Vuuren: [aico.van.vuuren@economics.gu.se](mailto:aico.van.vuuren@economics.gu.se)

Francis Vella: [francis.vella@georgetown.edu](mailto:francis.vella@georgetown.edu)

Franco Peracchi: [franco.peracchi@uniroma2.it](mailto:franco.peracchi@uniroma2.it)

We are grateful to useful comments and suggestions from the editor Chris Taber, two anonymous referees, Stéphane Bonhomme, Chinhui Juhn, Ivana Komunjer, Yona Rubinstein and Sami Stouli. Aico van Vuuren received financial support from Vetenskapsrådet, registration number 2020-02423.

developed estimators based on quantile regression. Start with a distribution regression model for  $F_{H^*|Z}$ . That is:

$$F_{H^*|Z}(h|x) = \Lambda(R(z)^T \gamma(h)),$$

where  $\Lambda$  is a known link function and  $R(z)$  is a  $d_R$ -dimensional vector of transformations of  $z$  with good approximating properties. We also assume a binary response model for the propensity score of selection:

$$\pi(z) = \Lambda(P(z)^T \pi_0),$$

where  $p(z)$  is a  $d_P$ -dimensional vector of transformations with good approximation properties such that:  $\Lambda(R(z)^T \gamma(\mu(z))) \approx 1 - \Lambda(P(z)^T \delta)$ .

Let  $\{Y_i D_i, H_i, D_i, Z_i\}_{i=1}^n$  be a random sample of  $(HY, H, D, Z)$ , where  $HY$  denotes that  $Y$  is only observed when  $D = 1$ . The proposed estimator consists of 2 steps:

1. Estimation of  $\pi(z)$  using binary regression in the entire sample:  $\hat{\pi}(z) = \Lambda(P(z)^T \hat{\delta})$  where:

$$\hat{\pi}_0 \in \arg \max_{p \in \mathbb{R}^{d_P}} \sum_{i=1}^n \left\{ D_i \log \Lambda(P(Z_i)^T p) + (1 - D_i) \log [1 - \Lambda(P(Z_i)^T p)] \right\}.$$

2. Estimation of  $F_{H^*|Z}$  by distribution regression with sample selection correction in the selected sample:  $\hat{F}_{H^*|Z}(h|z) = \Lambda(R(z)^T \hat{\gamma}(h))$  where:

$$\begin{aligned} \hat{\gamma}(h) \in \arg \max_{c \in \mathbb{R}^{d_R}} \sum_{i=1}^n D_i \left\{ 1(H_i \leq h) \log \left[ \Lambda(R(Z_i)^T c) + \hat{\pi}(Z_i) - 1 \right] \right. \\ \left. + 1(H_i > h) \log \left[ 1 - \Lambda(R(Z_i)^T c) \right] \right\}. \end{aligned}$$

## Web Appendix B: Proof of Lemma 2

$$\begin{aligned} \frac{1}{1-p} \int_p^1 G(y, x, v) dv &= \frac{1}{1-p} \int_p^1 \mathbb{P}(g(x, E) \leq y | V = v) dv \\ &= \int_p^1 \mathbb{P}(g(x, E) \leq y | V = v) dF_{V|V>p}(v) \\ &= \mathbb{P}(g(x, E) \leq y | V > \mu_h(z), Z = z, \mu_h(Z) = p) \\ &= \mathbb{P}(Y \leq y | H > h, X = x, \mu_h(Z) = p), \end{aligned}$$

The first equality is definition. The second equality uses  $V \sim U(0, 1)$  and the third equality uses independence of  $(E, V)$  and  $Z$ . The final equality uses the definitions of  $Y$  and  $H$  and is identified because  $(x, p) \in \mathcal{X}\mathcal{P}_K$ .

## Web Appendix C: Derivations of equation (12)

Adapting the representation of the distribution of the observed  $Y$  in Section 3.2 to the ordered selection rules yields

$$G_Y^s(y) = \frac{\int_{\mathcal{Z}^k} \int G(y, x, v) \mathbf{1}\{v > \mu_0(z)\} dv dF_Z(z)}{\int_{\mathcal{Z}^k} \int \mathbf{1}\{v > \mu_0(z)\} dv dF_Z(z)} = \frac{\sum_{h=1}^K \int_{\mathcal{Z}^k} \int G(y, x, v) \mathbf{1}\{\mu_{h-1}(z) < v \leq \mu_h(z)\} dv dF_Z(z)}{\int_{\mathcal{Z}^k} \int \mathbf{1}\{v > \mu_0(z)\} dv dF_Z(z)},$$

where the second equality uses that the interval  $(\mu_0(z), 1]$  is the union of the disjoint intervals  $(\mu_{h-1}(z), \mu_h(z)]$ ,  $h = 1, \dots, K$ .

## Web Appendix D: Details of the empirical implementation

This Appendix outlines the steps to estimate the model and calculate the counterfactual distributions under the three different selection mechanisms presented above. The data comprises  $(D_i^j, D_i^j Y_i^j, H_i^j, W_i^j, Z_i^j)$  where the notation is the same as above and recalling  $i = 1, \dots, n_j$  for  $j = \{t, k, r\}$  where  $t$  captures the year of structure,  $k$  captures the year of composition, and  $r$  captures the year of the selection process. When estimating the model parameters we take  $t, k$  and  $r$  for the same time period. The counterfactuals are based on different combinations of  $t, k$  and  $r$ .

### Web Appendix D.1: Model with censored selection

ALGORITHM 1. We implement the following steps to estimate the counterfactual distribution  $G_{Y_{(t,k,r)}}^s(y)$  for  $y \in \mathcal{Y} := \{y_1, y_2, \dots, y_{n_y}\}$ ;

#### 1. Probability of being censored:

Estimate a logit binary response model for the probability of censoring:

$$\pi(z) = \Lambda(P(z)^T \pi_0),$$

where  $\Lambda(\cdot)$  is the logistic function and where  $P(z)$  is a vector of transformations of  $z$  of dimension  $d_P$ . We estimate  $\pi^j(z) = \mathbb{P}(H^j > 0 | Z^j = z)$  for  $j = k, r$  via logit. That is:

$$\hat{\pi}_0^j = \arg \max_{p \in \mathbb{R}^{d_P}} \sum_{i=1}^{n_j} \left\{ D_i^j \log \Lambda(P(Z_i^j)^T p) + (1 - D_i^j) \log [1 - \Lambda(P(Z_i^j)^T p)] \right\}.$$

$\hat{\pi}^j(z) := \Lambda(P(z)^T \hat{\pi}_0^j)$  is the plugin estimator of  $\pi^j(z)$ .

## 2. Control function:

The control function is estimated as the conditional distribution of hours while accounting for the probability of being censored. This is defined as:

$$F_{H^*|Z}(h|z) = \Lambda(R(z)^T \gamma^j(h)),$$

where  $R(z)$  is a vector of transformations of  $z$  with dimension  $d_R$ . We estimate  $\gamma^j(h)$  for  $j = t, k$  via a series of logit regressions for all  $h = H_i^j$  in the selected sample for population  $j$ . That is:

$$\hat{\gamma}^j(h) = \arg \max_{c \in \mathbb{R}^{d_R}} \sum_{i=1}^{n_j} D_i^j \left\{ 1(H_i^j \leq h) \log \left[ \Lambda \left( R(Z_i^j)^T c \right) + \hat{\pi}^j(Z_i^j) - 1 \right] \right. \\ \left. + 1(H_i^j > h) \log \left[ 1 - \Lambda \left( R(Z_i^j)^T c \right) \right] \right\}.$$

Then:

$$\hat{V}_i^j = \Lambda \left( R(Z_i^j)^T \hat{\gamma}^j(H_i^j) \right),$$

is the plugin estimator of the control function  $V_i^j$ . The number of estimated  $\gamma^j$ 's corresponds to the number of unique values of  $H^j$ .

## 3. Conditional Wage Distribution:

We estimate the conditional distribution of wages as  $G(y, x, v) = \Lambda(Q(x, v)^T \beta(y))$ , where  $Q(x, v)$  is a  $d_Q$ -dimensional vector of transformations of  $(x, v)$ . We estimate  $\beta^t(y)$  via a series of logistic regressions. That is:

$$\hat{\beta}^t(y) = \arg \max_{b \in \mathbb{R}^{d_Q}} \sum_{i=1}^{n_t} D_i^t \left[ 1\{Y_i^t \leq y\} \log \Lambda(\hat{Q}_i^t b) + 1\{Y_i^t > y\} \log \Lambda(-\hat{Q}_i^t b) \right],$$

where  $\hat{Q}_i^t = Q(X_i^t, \hat{V}_i^t)$ . We employ the elements of  $X$ ,  $V$  and  $V^2$  as well as all elements of  $X$  interacted with  $V$ . The number of regressions equals the number of unique elements in  $\mathcal{Y}$ .

## 4. Counterfactual Wage Distribution:

We combine the parameters  $\hat{\beta}^t(y)$  with  $Z_i^{jT}$  and  $\hat{V}_i^j$  and use the logistic distribution to produce the counterfactual distributions  $G_{Y_{(t,k,r)}}^s(y)$ :

$$\hat{G}_{Y_{(t,k,r)}}^s(y) = \frac{1}{n_{kr}^s} \sum_{i=1}^{n_k} \Lambda(\hat{W}_i^{kT} \hat{\beta}^t(y)) 1\{\hat{V}_i^k > 1 - \hat{\pi}_r(Z_i^k)\},$$

where  $n_{kr}^s = \sum_{i=1}^n 1\{\hat{V}_i^k > 1 - \hat{\pi}_r(Z_i)\}$  is the number of observations for which the control function in the  $k$ -sample is higher than the cutoff value of the probability of not being in the selected sample. The summation is over all those observations for which the control function in the  $k$ -sample is higher than the cutoff value of the probability of not being in the selected sample. This is the sample analog of (5).

## Web Appendix D.2: Double selection mechanism

ALGORITHM 2. We implement the following steps to estimate the counterfactual distributions in the presence of two selection rules.

### 1. Probability of being censored:

Estimate two binary response models by logit explaining whether  $H$  and  $W$  are each censored:

$$\begin{aligned}\pi_H(z) &:= \mathbb{P}(H > 0 | Z = z) = \Lambda(P_H(z)^T \pi_{H,0}), \\ \pi_W(z) &:= \mathbb{P}(W > 0 | Z = z) = \Lambda(P_W(z)^T \pi_{W,0}),\end{aligned}$$

where  $P_H(z)$  and  $(P_W(z))$  are vectors of transformations of  $z$  with dimensions  $d_{P_H}$  and  $d_{P_W}$ . Define  $\hat{\pi}_H^j(z) := \Lambda(P_H(z)^T \hat{\pi}_{H,0})$  and  $\hat{\pi}_W^j(z) := \Lambda(P_W(z)^T \hat{\pi}_{W,0})$ , where  $\hat{\pi}_{H,0}$  is the estimator of  $\pi_{H,0}$  and  $\hat{\pi}_{W,0}$  is the estimator of  $\pi_{W,0}$ .

### 2. Control functions:

The control functions are estimated as the conditional distributions of hours and wages accounting for the probability of being censored. These are defined as:

$$\begin{aligned}F_{H(Z, V_H) | Z}(h | z) &= \Lambda \left( R_H(z)^T \gamma_H(h) \right), \\ F_{W(Z, V_W) | Z}(w | z) &= \Lambda \left( R_W(z)^T \gamma_W(w) \right),\end{aligned}$$

where  $R_H(z)$  and  $R_W(z)$  are vectors of transformations of  $z$  with dimensions  $d_{R_H}$  and  $d_{R_W}$ . We estimate each of these control functions via a series of logit regressions as is done for the single selection model for all  $h = H_i^j$  and  $w = W_i^j$  in the selected sample for population  $j$ . That is:

$$\begin{aligned}\hat{\gamma}_H^j(h) = \arg \max_{c \in \mathbb{R}^{d_{R_H}}} \sum_{i=1}^{n_j} 1(H_i^j > 0) \left\{ 1(H_i^j \leq h) \log \left[ \Lambda \left( R_H(Z_i^j)^T c \right) + \hat{\pi}_H^j(Z_i^j) - 1 \right] \right. \\ \left. + 1(H_i^j > h) \log \left[ 1 - \Lambda \left( R_H(Z_i^j)^T c \right) \right] \right\},\end{aligned}$$

and

$$\begin{aligned}\hat{\gamma}_W^j(w) = \arg \max_{c \in \mathbb{R}^{d_{R_W}}} \sum_{i=1}^{n_j} 1(W_i^j > 0) \left\{ 1(W_i^j \leq w) \log \left[ \Lambda \left( R_W(Z_i^j)^T c \right) + \hat{\pi}_W^j(Z_i^j) - 1 \right] \right. \\ \left. + 1(W_i^j > w) \log \left[ 1 - \Lambda \left( R_W(Z_i^j)^T c \right) \right] \right\}.\end{aligned}$$

The estimated control functions are:

$$\begin{aligned}\hat{V}_{H,i}^j &= \Lambda \left( R_H(Z_i^j)^T \hat{\gamma}_H^j(H_i^j) \right), \\ \hat{V}_{W,i}^j &= \Lambda \left( R_W(Z_i^j)^T \hat{\gamma}_W^j(W_i^j) \right).\end{aligned}$$

### 3. Conditional Wage Distribution:

We impose  $G(y, x, v_H, v_W) = \Lambda(Q(x, v_H, v_W)^T \beta(y))$ , where  $Q(x, v_H, v_W)$  is a  $d_Q$ -dimensional vector of transformations of  $(x, v_H, v_W)$ . We estimate  $\beta^t(y)$  via a series of logit regressions:

$$\hat{\beta}^t(y) = \arg \max_{b \in \mathbb{R}^{d_Q}} \sum_{i=1}^{n_t} 1(H_i^t > 0) 1(W_i^t > 0) \left[ 1\{Y_i^t \leq y\} \log \Lambda(\hat{Q}_i^{tT} b) + 1\{Y_i^t > y\} \log \Lambda(-\hat{Q}_i^{tT} b) \right],$$

where  $\hat{Q}_i^t = Q(X_i^t, \hat{V}_i^t)$ .

### 4. Counterfactual Wage Distributions:

We estimate the counterfactual distribution  $G_{Y_{(t,k,r)}}^s(y)$  by:

$$\hat{G}_{Y_{(t,k,r)}}^s(y) = \frac{1}{n_{kr}^s} \sum_{i=1}^{n_k} \Lambda(\hat{Q}_i^{kT} \hat{\beta}^t(y)) 1\{\hat{V}_{H,i}^k > 1 - \hat{\pi}_H^r(Z_i^k)\} 1\{\hat{V}_{W,i}^k > 1 - \hat{\pi}_W^r(Z_i^k)\},$$

where  $n_{kr}^s = \sum_{i=1}^n 1\{\hat{V}_{H,i}^k > 1 - \hat{\pi}_H^r(Z_i)\} 1\{\hat{V}_{W,i}^k > 1 - \hat{\pi}_W^r(Z_i)\}$ .

## Web Appendix D.3: Model with ordered selection

ALGORITHM 3. We implement the following steps to estimate the counterfactual distributions in the presence of an ordered selection rule.

#### 1. Control functions:

We estimate the control function by imposing the following probabilities for the categorical hours variable  $H$ :

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

for  $0 \leq h < K$  and where  $P(z)$  is a vector of transformations with dimension  $d_P$ . For  $j = t, k, r$ , we estimate  $\gamma_h; h = 0, \dots, K - 1$  by solving the following maximization problem:

$$\begin{aligned} \hat{\gamma}^j = \arg \max_{c \in \mathbb{R}^{(K-1) \times d_P}} & \sum_{i=1}^n \left\{ 1(H_i = 0) \log \Lambda \left( P(Z_i^k)^T c_0 \right) \right. \\ & + \sum_{h=1}^{K-1} 1(H_i = h) \left[ \log \Lambda \left( P(Z_i^k)^T c_h \right) - \log \Lambda \left( P(Z_i^k)^T c_{h-1} \right) \right] \\ & \left. + 1(H_i = K) \left[ 1 - \log \Lambda \left( P(Z_i^k)^T c_{K-1} \right) \right] \right\}, \end{aligned}$$

where  $\gamma^j = (\gamma_0^j, \gamma_1^j, \dots, \gamma_{K-1}^j)$  and  $c = (c_1, c_2, \dots, c_{K-1})$ . This is not a standard ordered logit model as the value of  $\gamma_h$  varies by category. The control functions are

estimated as:

$$\widehat{\mu}_{i,h}^j = \Lambda \left( P(Z_i^j)^T \widehat{\gamma}_h^j \right),$$

for  $0 \leq h < K$ .

## 2. Conditional Wage Distribution:

We impose:

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

where  $Q(x, \mu_h, \mu_{h-1})$  is a  $d_Q$ -dimensional vector of transformations of  $(x, \mu_h, \mu_{h-1})$ . We employ the elements of  $X$  as well as the full set of  $\widehat{\mu}_{i,h}$ ;  $h = 1, \dots, K$ , interacted with the relevant dummy variables of  $H$  and the elements of  $X$ . We estimate  $\beta$  via series of logit regressions. That is:

$$\widehat{\beta}^t(y) = \arg \max_{b \in \mathbb{R}^{d_Q}} \sum_{i=1}^{n_t} D_i^t \left[ 1\{Y_i^t \leq y\} \log \Lambda(\widehat{Q}_i^t b) + 1\{Y_i^t > y\} \log \Lambda(-\widehat{Q}_i^t b) \right].$$

where  $\widehat{Q}_i^t = Q(X_i^t, \widehat{\mu}_{i,h-1}^t, \widehat{\mu}_{i,h}^t)$ .

## 3. Counterfactual Wage Distribution

The counterfactual wage distributions  $G_{Y_{(t,k,r)}}^s(y)$  are estimated by:

$$\frac{\sum_{i=1}^{n_k} \Lambda(\widehat{Q}_i^k b^t(y)) (\widehat{\mu}_{1,i}^k - \widehat{\mu}_{0,i}^r)}{\sum_{i=1}^n (1 - \widehat{\mu}_{0,i}^r)} + \sum_{h=2}^K \frac{\sum_{i=1}^{n_k} \Lambda(\widehat{Q}_i^h b^t(y)) (\widehat{\mu}_{h,i}^k - \widehat{\mu}_{h-1,i}^k)}{\sum_{i=1}^n (1 - \widehat{\mu}_{0,i}^r)}.$$

This estimator is the sample analog of (13).

## Web Appendix E: Figures

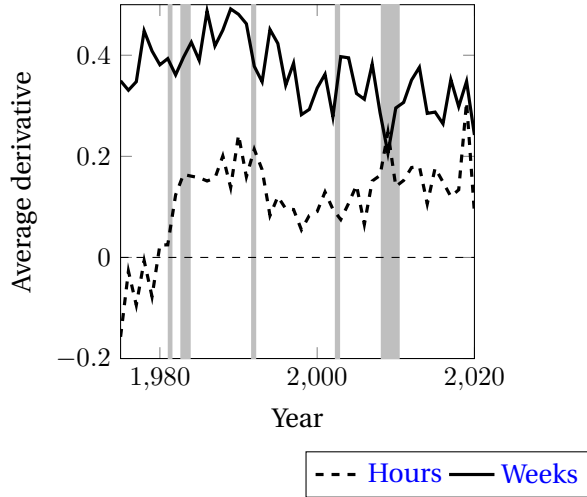


FIGURE WEB-1. Average derivative

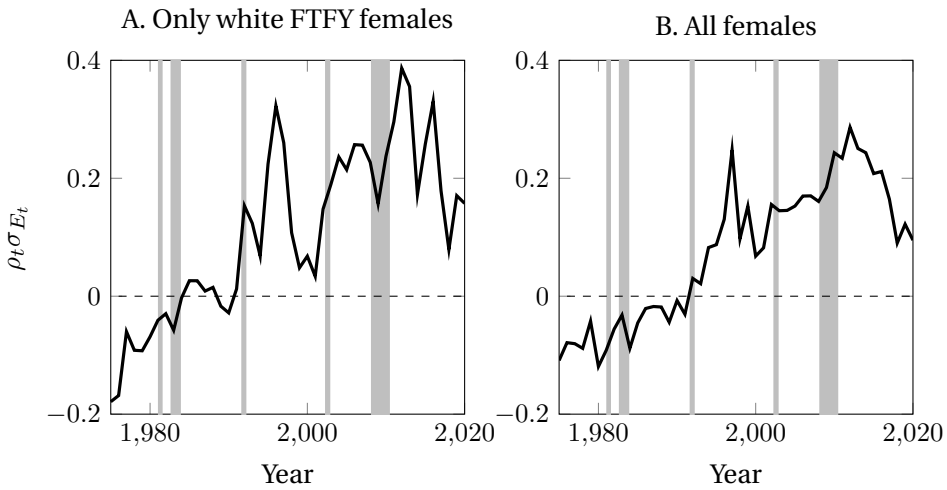


FIGURE WEB-2. Estimation of  $\rho_t \sigma_{E_t}$  over time.



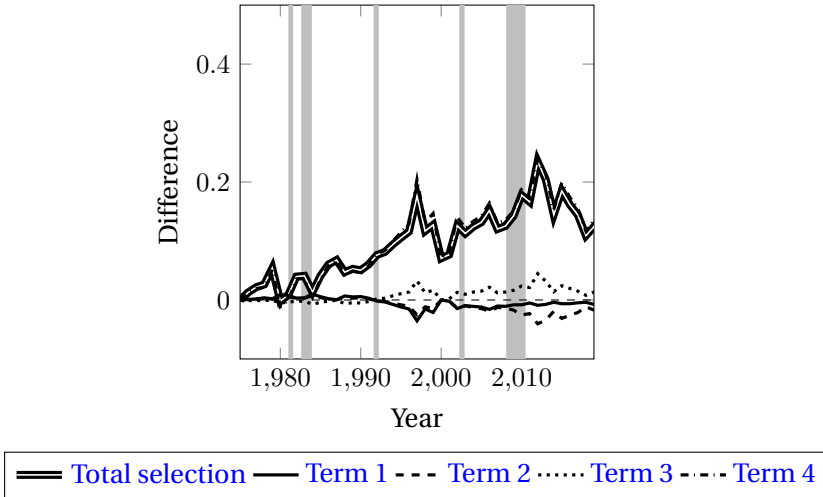


FIGURE WEB-3. Selection effect using the Heckman-selection model and the 4 parts of the selection effect are represented by (18); total sample is full population between 24 and 65 and selected population is all workers working a positive number of hours.

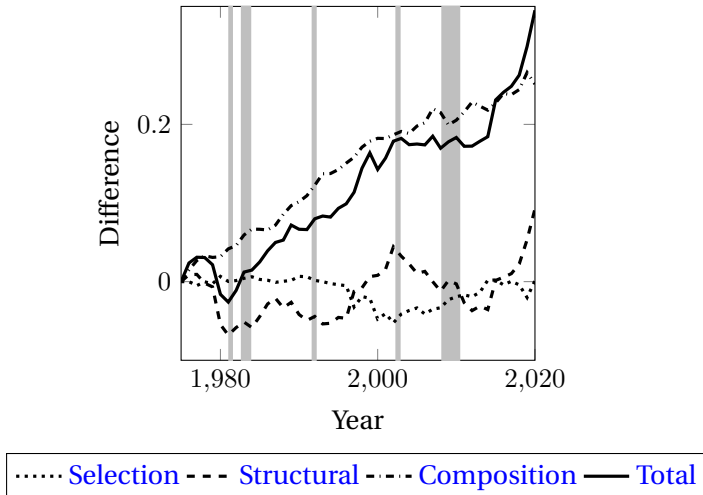


FIGURE WEB-4. Computation of the selection, composition and structural effect based on equation (19)-(21); total sample is full population between 24 and 65 and selected population is all workers working a positive number of hours.

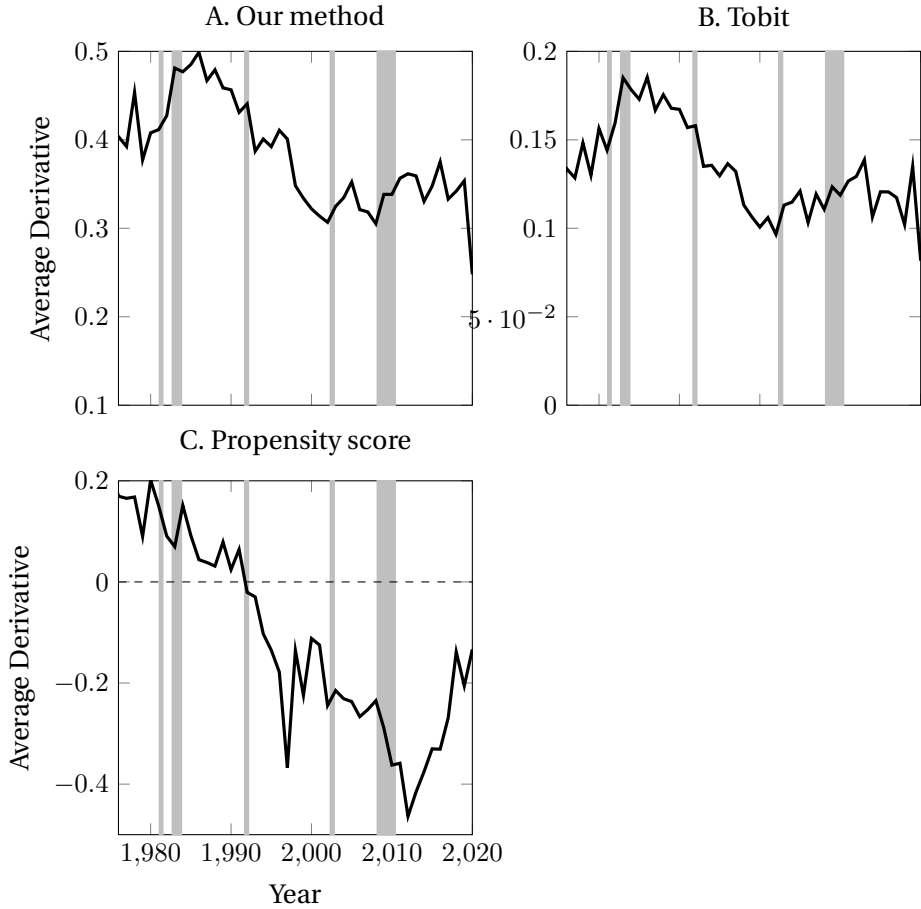


FIGURE WEB-5. Average derivative of the wage. A. is the derivative of our control function, B. is the average derivative using derivatives of a Tobit, C. uses a non-separable sample selection model.