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

(*Quantitative Economics*, Vol. 14, No. 2, May 2023, 571–607)

Iván Fernández-Val
Department of Economics, Boston University

Aico van Vuuren
Department of Economics, Gothenburg University and Department of Economics, University of Groningen

Francis Vella
Department of Economics, Georgetown University

Franco Peracchi
Department of Economics, University of Rome Tor Vergata

This Supplemantary Appendix contains the description of the distribution regression estimator of the control function for models with unknown censoring point, proof of Lemma 2, derivation of equation (12), more details on the empirical implementation, and some figures omitted from the main text.

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 $F_{H^*\mid X,Z}$ based on distribution regression. This is an alternative to Buchinsky and Hahn (1998) and Chernozhukov and Hong (2002), which developed estimators based on quantile regression. Start with a distribution regression model for $F_{H^*\mid Z}$. That is,

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

where Λ 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

Iván Fernández-Val: ivanf@bu.edu
Aico van Vuuren: aico.van.vuuren@economics.gu.se
Francis Vella: francis.vella@georgetown.edu
Franco Peracchi: franco.peracchi@uniroma2.it
We are grateful to useful comments and suggestions from 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.

 https://doi.org/10.3982/QE1777

the propensity score of selection:

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

where $P(z)$ is a d_P-dimensional vector of transformations with good approximation properties such that $\Lambda(R(z)^T \gamma(\mu(z))) \approx 1 - \Lambda(P(z)^T \pi_0)$ where $\mu(z)$ is the unknown censoring point.

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: $\widehat{\pi}(z) = \Lambda(P(z)^T \widehat{\pi}_0)$ where

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

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

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

Web Appendix B: Proof of Lemma 2

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

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 $(x, p) \in \mathcal{XP}_K$.

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

$$G_Y^s(y) = \frac{\int_{\mathcal{Z}^k} \int G(y, x, v) 1\big\{v > \mu_0(z)\big\}\, dv\, dF_Z(z)}{\int_{\mathcal{Z}^k} \int 1\big\{v > \mu_0(z)\big\}\, dv\, dF_Z(z)}$$

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

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.

D.1 *Model with censored selection*

Algorithm 1. We implement the following steps to estimate the counterfactual distribution $G^s_{Y_{\langle t,k,r \rangle}}(y)$ 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:

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

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

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

   $\widehat{\pi}_0^j(z) := \Lambda(P(z)^T \widehat{\pi}^j)$ is the plug-in 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

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

   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,

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

Then

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

is the plug-in estimator of the control function V_i^j. The number of estimated γ^js corresponds to the number of observed 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 logisitic regressions. That is,

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

   where $\widehat{Q}_i^t = Q(X_i^t, \widehat{V}_i^t)$. 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 $\widehat{\beta}^t(y)$ with $Z_i^{j^T}$ and $\widehat{V}_i^j$ and use the logistic distribution to produce the counterfactual distributions $G^s_{Y_{\langle t,k,r\rangle}}(y)$:

$$\widehat{G}^s_{Y_{\langle t,k,r\rangle}}(y) = \frac{1}{n^s_{kr}} \sum_{i=1}^{n_k} \Lambda(\widehat{W}_i^{k^T}\widehat{\beta}^t(y)) 1\{\widehat{V}_i^k > 1 - \widehat{\pi}_r(Z_i^k)\},$$

   where $n^s_{kr} = \sum_{i=1}^n 1\{\widehat{V}_i^k > 1 - \widehat{\pi}_r(Z_i)\}$ is the number of observations for which the control function in the k-sample is higher than the cut-off 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).

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:

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

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

$$F_{H(Z,V_H)\mid Z}(h \mid z) = \Lambda\big(R_H(z)^T \gamma_H(h)\big),$$
$$F_{W(Z,V_W)\mid Z}(w \mid z) = \Lambda\big(R_W(z)^T \gamma_W(w)\big),$$

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 $h = H^j_i$ and $w = W^j_i$ in the selected sample for population j. That is,

$$\begin{aligned}\widehat{\gamma}^j_H(h) = \arg\max_{c \in \mathbb{R}^{d_{R_H}}} \sum_{i=1}^{n_j} 1\big(H^j_i > 0\big)\big\{1\big(H^j_i \le h\big)\log\big[\Lambda\big(R_H(Z^j_i)^T c\big) + \widehat{\pi}^j_H(Z^j_i) - 1\big] \\ + 1\big(H^j_i > h\big)\log\big[1 - \Lambda\big(R_H(Z^j_i)^T c\big)\big]\big\},\end{aligned}$$

and

$$\begin{aligned}\widehat{\gamma}^j_W(w) = \arg\max_{c \in \mathbb{R}^{d_{R_W}}} \sum_{i=1}^{n_j} 1\big(W^j_i > 0\big)\big\{1\big(W^j_i \le w\big)\log\big[\Lambda\big(R_W(Z^j_i)^T c\big) + \widehat{\pi}^j_W(Z^j_i) - 1\big] \\ + 1\big(W^j_i > w\big)\log\big[1 - \Lambda\big(R_W(Z^j_i)^T c\big)\big]\big\}.\end{aligned}$$

The estimated control functions are

$$\widehat{V}^j_{H,i} = \Lambda\big(R_H(Z^j_i)^T \widehat{\gamma}^j_H(H^j_i)\big),$$
$$\widehat{V}^j_{W,i} = \Lambda\big(R_W(Z^j_i)^T \widehat{\gamma}^j_W(W^j_i)\big).$$

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:

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

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

4. **Counterfactual Wage Distributions:**
   We estimate the counterfactual distribution $G^s_{Y_{\langle t,k,r \rangle}}(y)$ by

$$\widehat{G}^s_{Y_{\langle t,k,r \rangle}}(y) = \frac{1}{n^s_{kr}} \sum_{i=1}^{n_k} \Lambda\left(\widehat{Q}_i^{k^T} \widehat{\beta}^t(y)\right) 1\left\{\widehat{V}^k_{H,i} > 1 - \widehat{\pi}^r_H\left(Z_i^k\right)\right\} 1\left\{\widehat{V}^k_{W,i} > 1 - \widehat{\pi}^r_W\left(Z_i^k\right)\right\},$$

where $n^s_{kr} = \sum_{i=1}^n 1\{\widehat{V}^k_{H,i} > 1 - \widehat{\pi}^r_H(Z_i)\} 1\{\widehat{V}^k_{W,i} > 1 - \widehat{\pi}^r_W(Z_i)\}$.

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:

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

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 γ_h; $h = 0, \ldots, K-1$ by solving the following maximization problem:

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

where $\gamma^j = (\gamma_0^j, \gamma_2^j, \ldots, \gamma_{K-1}^j)$ and $c = (c_1, c_2, \ldots, c_{K-1})$. This is not a standard ordered logit model as the value of γ_h varies by category. The control functions are estimated as

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

for $0 \leq h < K$.

2. **Conditional Wage Distribution:**

   We impose

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

   where $Q(x,\mu_h,\mu_{h-1})$ is a d_Q-dimensional vector of transformations of (x,μ_h,μ_{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 β via series of logit regressions. That is,

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

   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 $G^s_{Y_{\langle t,k,r\rangle}}(y)$ are estimated by

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

   This estimator is the sample analog of (13).

Web Appendix E: Figures

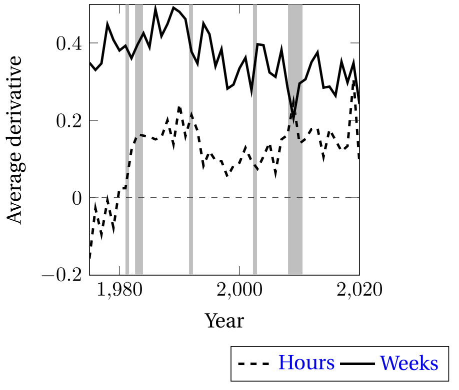

Figure Web-1. Average derivative.

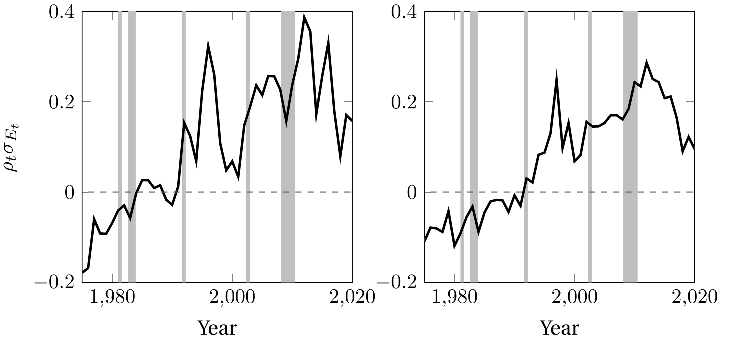

Figure Web-2. Estimation of $\rho_t \sigma_{E_t}$ over time.

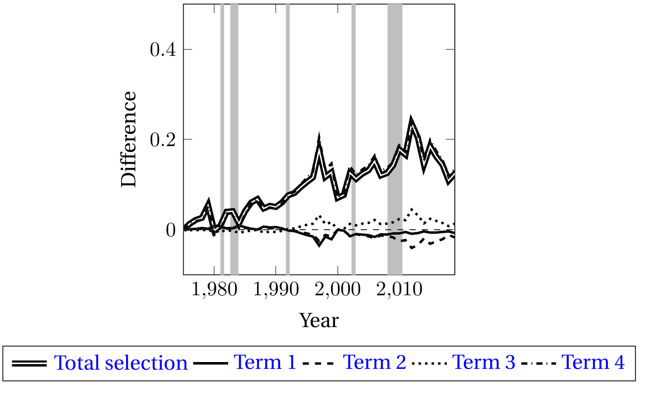

Figure Web-3. Selection effect using the Heckman-selection model and the 4 parts of the selection effect are represented by (18); total sample is full population between 24 and 65 and selected population is all workers working a positive number of hours.

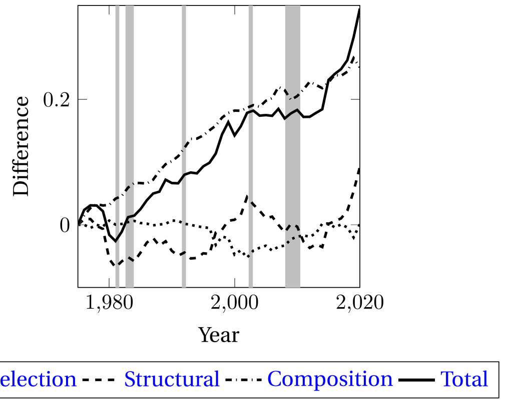

FIGURE WEB-4. Computation of the selection, composition and structural effect based on equation (19)–(21); total sample is full population between 24 and 65 and selected population is all workers working a positive number of hours.

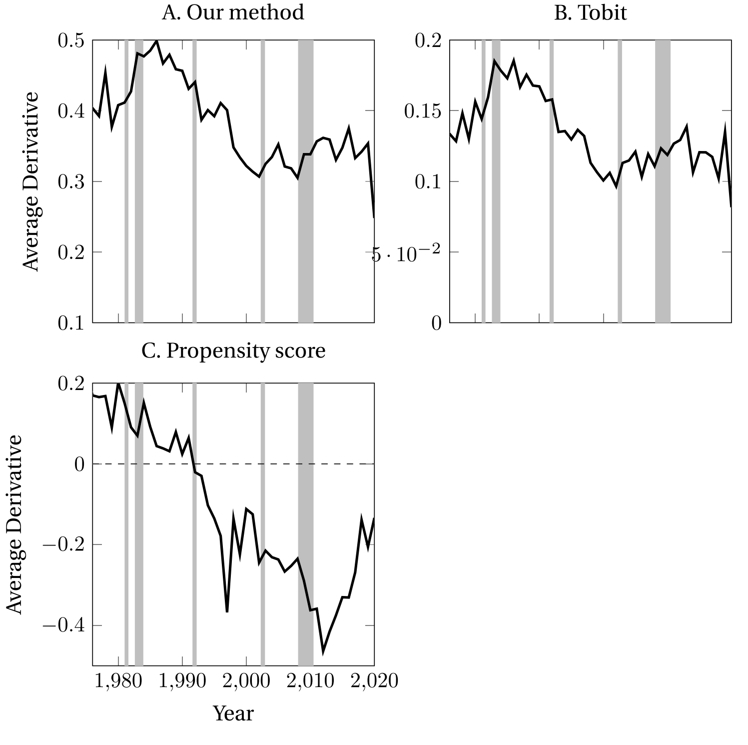

FIGURE WEB-5. Average derivative of the wage. (A) is the derivative of our control function, (B) is the average derivative using derivatives of a Tobit, (C) uses a nonseparable sample selection model.

References

Buchinsky, Moshe and Jinyong Hahn (1998), "An alternative estimator for the censored quantile regression model." *Econometrica*, 66, 653–671. [1]

Chernozhukov, Victor and Han Hong (2002), "Three-step censored quantile regression and extramarital affairs." *Journal of the American Statistical Association*, 97, 872–882. [1]

Co-editor Christopher Taber handled this manuscript.

Manuscript received 23 November, 2020; final version accepted 11 August, 2022; available online 30 August, 2022.