Supplement to "Sensitivity to missing data assumptions: Theory and an evaluation of the U.S. wage structure"

(*Quantitative Economics*, Vol. 4, No. 2, July 2013, 231–267)

Patrick Kline
Department of Economics, University of California, Berkeley and NBER

Andres Santos
Department of Economics, University of California, San Diego

Appendix A: The bivariate normal selection model and KS

To develop intuition for our metric $\mathcal{S}(F)$ of deviations from missing at random, we provide here a mapping between the parameters of a standard bivariate selection model, the resulting CDFs of observed and missing outcomes, and the implied values of $\mathcal{S}(F)$. Using the notation of Section 2, our data generating process (DGP) of interest is

$$(Y_i, v_i) \sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \qquad D_i = 1\{\mu + v_i > 0\}. \tag{A.1}$$

In this model, the parameter ρ indexes the degree of nonignorable selection in the outcome variable Y_i. We chose $\mu = 0.6745$ to ensure a missing fraction of 25%, which is approximately the degree of missingness found in our analysis of earnings data in the U.S. Census. We computed the distributions of missing and observed outcomes for various values of ρ by simulation, some of which are plotted in Figures A.1 and A.2. The resulting values of $\mathcal{S}(F)$, which correspond to the maximum vertical distance between the observed and missing CDFs across all points of evaluation, are given in Table A.1.

Table A.1. $\mathcal{S}(F)$ as a function of ρ.

ρ	$\mathcal{S}(F)$	ρ	$\mathcal{S}(F)$	ρ	$\mathcal{S}(F)$
0.05	0.0337	0.35	0.2433	0.65	0.4757
0.10	0.0672	0.40	0.2778	0.70	0.5165
0.15	0.1017	0.45	0.3138	0.75	0.5641
0.20	0.1355	0.50	0.3520	0.80	0.6158
0.25	0.1721	0.55	0.3892	0.85	0.6717
0.30	0.2069	0.60	0.4304	0.90	0.7377

Patrick Kline: pkline@econ.berkeley.edu
Andres Santos: a2santos@ucsd.edu

DOI: 10.3982/QE176

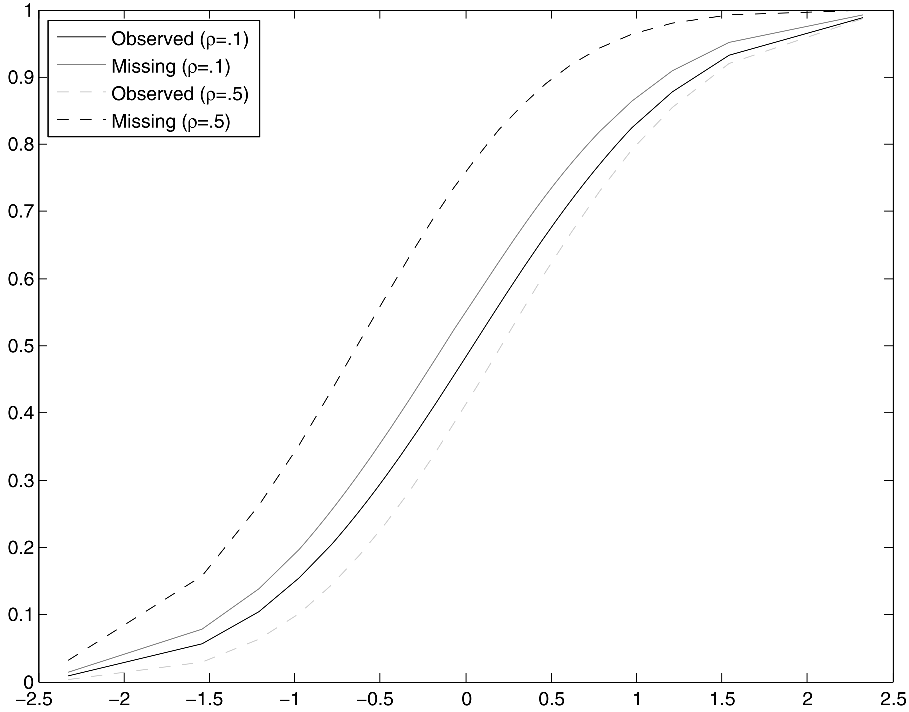

FIGURE A.1. Missing and observed outcome CDFs.

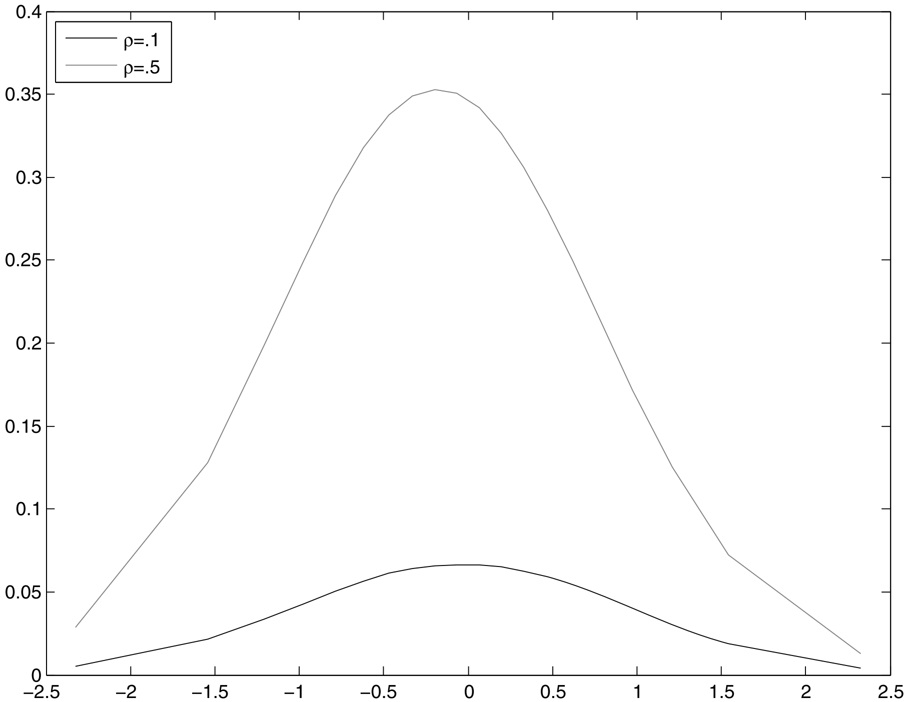

FIGURE A.2. Vertical distance between CDFs.

Appendix B: Data details

Census data

Our analysis of Decennial Census data uses 1% unweighted extracts obtained from the Minnesota IPUMS website http://usa.ipums.org/ on August 28th, 2009. Our extract contained all native born black and white men ages 40–49. We drop all men with less than 5 years of schooling and recode the schooling variable according to the scheme described by Angrist, Chernozhukov, and Fernádez-Val (2006) in their online appendix available at http://econ-www.mit.edu/files/385. Our IPUMS extract, along with Stata code used to impose our sample restrictions, is available online at http://qeconomics.org/supp/176/code_and_data.zip.

Like Angrist, Chernozhukov, and Fernádez-Val (2006), we drop from the sample all individuals with allocated age or education information and all individuals known not to have worked or generated earnings. We also drop observations in small demographic cells. Table B.2 lists the effects of these drops on the sample sizes by Census year.

Following Angrist, Chernozhukov, and Fernádez-Val (2006) we choose as our wage concept log average weekly earnings—the log of total annual wage and salary income divided by annual weeks worked. Earnings in all years are converted to 1989 dollars using the Personal Consumption Expenditure (PCE) price index. We recode to missing all weekly earnings observations with allocated earnings or weeks worked.

CPS data

For the analysis in Section 5.2, we used ICPSR (Inter-university Consortium for Political and Social Research) archive 7616, "Current Population Survey, 1973, and Social Security Records: Exact Match Data." We extract from this file a sample of white and black men ages 25–54 with 6 or more years of schooling who reported working at least one week in the last year. We then drop from the sample individuals who are self-employed and those who work in industries or occupations identified by Bound and Krueger (1991) as likely to receive tips that are underreported to the IRS.

Annual IRS wage and salary earnings are top-coded at $50,000 dollars. There are also a small fraction of observations with very low IRS earnings below $1,000. We drop observations that fall into either group. We also drop observations with allocated weeks worked. Finally we drop observations that fall into demographic cells with less than 50

Table B.2. Census sample sizes by year after imposing restrictions.

	1980	1990	2000
Native born black and white men ages 40–49			
w/$\geq$5 years of schooling	97,900	131,667	168,909
Drop observations w/imputed age	96,403	130,806	165,505
Drop observations w/imputed education	90,064	124,161	155,158
Drop observations w/unallocated (earnings or weeks worked) = 0	80,800	111,366	131,711
Drop observations in cells w/<20 observations	80,128	111,070	131,265

TABLE B.3. CPS sample sizes after imposing restrictions.

	1980
Black and white men ages 25–54 w/≥5 years of schooling and one or more weeks worked	19,693
Drop self-employed	17,665
Drop bound-Krueger industries/occupations	17,138
Drop top-coded IRS earnings and outliers	15,632
Drop observations w/allocated weeks worked	15,355
Drop cells w/<50 observations	15,027

observations. An itemization of the effect of these decisions on sample size is provided by Table B.3.

Weeks worked are reported categorically in the 1973 CPS. We code unallocated weeks responses to the midpoint of their interval. Our average weekly earnings measure is constructed by dividing IRS wage and salary earnings by the recoded weeks worked variable.

APPENDIX C: IMPLEMENTATION DETAILS

We outline here the implementation of our estimation and inference procedure in the Decennial Census data. The MATLAB code employed is available online at http://qeconomics.org/supp/176/code_and_data.zip.

Our estimation and inference routine consists of three distinct procedures: (i) Obtaining point estimates, (ii) obtaining bootstrap estimates, and (iii) constructing confidence regions from such estimates. Below we outline in detail the algorithms employed in each procedure.

(i) *Point Estimates*: We examine a grid of quantiles, denoted $\mathcal{T}$, with lower and upper limits of $\underline{\tau}$ and $\bar{\tau}$, for example $\mathcal{T} = \{0.1, 0.15, \ldots, 0.85, 0.9\}$. For each τ in this grid, we let

$$K_u(\tau) \equiv \min\left\{\frac{\min\{\tau, (1-\tau)\}}{\max_{x\in\mathcal{X}}(1-\hat{p}(x))}, \frac{\min\{\tau, (1-\tau)\}}{\max_{x\in\mathcal{X}} \hat{p}(x)}, 0.3\right\} - 0.001, \tag{C.1}$$

and examine for each $\tau \in \mathcal{T}$, a grid of restrictions k, denoted $\mathcal{K}(\tau)$, with a lower bound of 0 and an upper bound $K_u(\tau)$, for example, $\mathcal{K}(\tau) = \{0, 0.01, \ldots, \lfloor\frac{K_u(\tau)}{0.01}\rfloor \times 0.01\}$. These grids approximate

$$\hat{\mathcal{B}} \equiv \{(\tau, k) \in [0,1]^2 : \underline{\tau} \leq \tau \leq \bar{\tau} \text{ and } 0 \leq k \leq K_u(\tau)\}, \tag{C.2}$$

which is, with probability tending to 1, a subset of $\mathcal{B}_\zeta$ for some ζ such that $\mathcal{B}_\zeta \neq \emptyset$. For each pair (τ, k) in our constructed grid, we then perform the following operations:

STEP 1. For each $x \in \mathcal{X}$, we find $\hat{q}_L(\tau, k|x)$ and $\hat{q}_U(\tau, k|x)$, which respectively are just the $\tau - k(1-\hat{p}(x))$ and $\tau + k(1+\hat{p}(x))$ quantiles of observed Y for the demographic group with $X = x$.

STEP 2. We obtain $\hat{\pi}_L(\tau, k)$ and $\hat{\pi}_U(\tau, k)$ by solving the linear programming problems defined in equations (36) and (37), employing the `linprog` routine in MATLAB.

(ii) *Bootstrap Estimates*: We generate an i.i.d. sample $\{W_i\}_{i=1}^n$ with W exponentially distributed and $E[W] = \text{Var}(W) = 1$. For each (τ, k) in the grid employed to find point estimates, we then perform the following operations:

STEP 1. For each $x \in \mathcal{X}$, we find $\tilde{q}_L(\tau, k|x)$ and $\tilde{q}_U(\tau, k|x)$. Computationally, they respectively equal the $\tau - k(1 - \tilde{p}(x))$ and $\tau + k(1 - \tilde{p}(x))$ weighted quantiles of observed Y for the group $X = x$, where each observation receives weight $W_i/(\sum W_i 1\{D_i = 1, X_i = x\})$ rather than $1/(\sum_i 1\{D_i = 1, X_i = x\})$.

STEP 2. We obtain $\tilde{\pi}_L(\tau, k)$ and $\tilde{\pi}_U(\tau, k)$ by solving the linear programming problems defined in equations (53) and (54) by employing the `linprog` routine in MATLAB.

(iii) *Confidence Regions*: Throughout we set $\omega_L(\tau, k) = \omega_U(\tau, k) = \omega(\tau)$, where $\omega(\tau) \equiv \phi(\Phi^{-1}(\tau))^{-1/2}$, with $\phi(\cdot)$ and $\Phi(\cdot)$ equal to the standard normal density and CDF. In computing confidence regions, we employ the point estimates $(\hat{\pi}_L^t, \hat{\pi}_U^t)$ for $t \in \{80, 90, 00\}$ from (i), and 1,000 bootstrap estimates $\{(\tilde{\pi}_{b,L}^t, \tilde{\pi}_{b,U}^t)\}_{b=1}^{1{,}000}$ computed according to (ii) based on 1,000 independent i.i.d. samples $\{W_i\}_{i=1}^n$.

The specifics underlying the computation of each figure's confidence regions are as follows:

FIGURE 4. For this figure, we compute $(\hat{q}_L(\tau, k|x), \hat{q}_U(\tau, k|x))$ evaluated at $(\tau, k) = (0.5, 1)$ for all $x \in \mathcal{X}$. Employing the bootstrap analogues $(\tilde{q}_L(0.5, 1|x), \tilde{q}_U(0.5, 1|x))$, we obtain estimates $\hat{\sigma}_L$ and $\hat{\sigma}_U$ of the asymptotic variances of $\hat{q}_L(0.5, 1|x)$ and $\hat{q}_U(0.5, 1|x)$, and construct

$$\left[\hat{q}_L(0.5, 1|x) - \frac{z_{1-\alpha}\hat{\sigma}_L}{\sqrt{n}}, \hat{q}_U(0.5, 1|x) + \frac{z_{1-\alpha}\hat{\sigma}_U}{\sqrt{n}}\right] \tag{C.3}$$

for all $x \in \mathcal{X}$, where $z_{1-\alpha}$ is the $1 - \alpha$ quantile of a standard normal. By independence, the product of the intervals (C.3) evaluated at $\alpha = 0.95^{1/227}$ is a nonparametric confidence region with asymptotic probability of covering $q(\tau|x)$ of at least 0.95 (Imbens and Manski (2004)). Employing `linprog` in MATLAB, we obtain the bounds for the "parametric set" by maximizing/minimizing the coefficient on schooling subject to the constraint that the Mincer specification lies within (C.3) for all $x \in \mathcal{X}$.

FIGURE 5. The nonparametric confidence region was obtained as in Figure 4, but employing $k = 0.05$ instead. The bounds on the set of BLPs were obtained by solving a linear programming problem as defined in equations (24) and (25), but employing the endpoints of the interval defined in (C.3) in place of $q_L(0.5, 1|x)$ and $q_U(0.5, 1|x)$; here λ

equals 1 for the coordinate corresponding to the coefficient on education, and equals 0 elsewhere.

FIGURE 6. We employ a grid for τ equal to $\mathcal{T} = \{0.1, 0.15, \dots, 0.85, 0.9\}$. For each $t \in \{80, 90, 00\}$, we compute the $1-\alpha$ quantile across bootstrap samples of

$$\max_{\tau \in \{0.1, 0.15, \dots, 0.85, 0.9\}} \max\left\{\frac{\tilde{\pi}_L^t(\tau, 0) - \hat{\pi}_L^t(\tau, 0)}{\omega(\tau)}, \frac{\hat{\pi}_U^t(\tau, 0) - \tilde{\pi}_U^t(\tau, 0)}{\omega(\tau)}\right\}, \tag{C.4}$$

which we denote by $\tilde{r}_{1-\alpha}^t(0)$. For each $t \in \{80, 90, 00\}$ the two-sided uniform confidence region is then given by $[\hat{\pi}_L^t(\tau, 0) - \tilde{r}_{1-\alpha}^t(0)\omega(\tau), \hat{\pi}_U^t(\tau, 0) + \tilde{r}_{1-\alpha}^t(0)\omega(\tau)]$, where for τ outside our grid $\{0.1, 0.15, \dots, 0.85, 0.9\}$, we obtain a number by linear interpolation.

FIGURE 7. The procedure is identical to Figure 6, except k is set at 0.05 instead of at 0.

FIGURE 8. For $k_s = 0.05$, we compute the $1-\alpha$ quantile across bootstrap samples of

$$\max_{\tau \in \{0.2, 0.25, \dots, 0.75, 0.8\}} \frac{\hat{\pi}_U^{80}(\tau, k_s) - \tilde{\pi}_U^{80}(\tau, k_s)}{\omega(\tau)}, \tag{C.5}$$

which we denote by $\tilde{r}_{1-\alpha}^{80}(k_s)$. Similarly, we find the $1-\alpha$ quantile across bootstrap samples of

$$\max_{\tau \in \{0.2, 0.25, \dots, 0.75, 0.8\}} \frac{\tilde{\pi}_L^{90}(\tau, k_s) - \hat{\pi}_L^{90}(\tau, k_s)}{\omega(\tau)}, \tag{C.6}$$

which we denote by $\tilde{r}_{1-\alpha}^{90}(k_s)$. Unlike in Figures 6 and 7, we employ a shorter grid for τ, as the bounds corresponding to extreme quantiles become unbounded for large k. Next, we examine whether

$$\begin{aligned} \min_{\tau \in \{0.2, 0.25, \dots, 0.75, 0.8\}} &\big\{\big(\hat{\pi}_U^{80}(\tau, k_s) + \tilde{r}_{1-\alpha}^{80}(k_s)\omega(\tau)\big) \\ &- \big(\hat{\pi}_L^{90}(\tau, k_s) - \tilde{r}_{1-\alpha}^{90}(k_s)\omega(\tau)\big)\big\} \geq 0. \end{aligned} \tag{C.7}$$

If (C.7) holds, we set $k_0^* = k_s$; otherwise repeat (C.5)–(C.7) with $k_s + 0.005$. Hence, $k_0^* = 0.175$ was the smallest k (under steps of size 0.005) for which the upper confidence interval for $\pi_U^{80}(\tau, k_0^*)$ laid above the lower confidence interval for $\pi_L^{90}(\tau, k_0^*)$ for all $\tau \in \{0.2, 0.25, \dots, 0.75, 0.8\}$.

FIGURE 9. For this figure, we employ a grid of size 0.01 for τ, for instance, $\mathcal{T} = \{0.1, 0.11, \dots, 0.89, 0.9\}$, and compute $K_u(\tau)$ (as in (C.1)) using the 1990 Decennial Census. In turn, for each τ, we employ a k-grid $\mathcal{K}(\tau) = \{0, 0.001, \dots, \lfloor \frac{K_u(\tau)}{0.001} \rfloor \times 0.001\}$, and obtain $\hat{\pi}_U^{80}(\tau, k)$ and $\hat{\pi}_L^{80}(\tau, k)$ at each (τ, k) pair. Finally, for each τ in our grid, we let $k_I(\tau)$ denote the smallest value of k in its grid such that

$$\hat{\pi}_U^{80}(\tau, k) \geq \hat{\pi}_L^{90}(\tau, k) \tag{C.8}$$

and define $\hat{\kappa}(\tau) \equiv (\hat{\pi}_U^{80}(\tau, k_I(\tau)) + \hat{\pi}_L^{90}(\tau, k_I(\tau)))/2$, which constitutes our estimate of the "breakdown curve." To obtain a confidence region, we compute the $1-\alpha$ quantile across bootstrap samples of

$$\max_{\tau \in \{0.1, 0,15, \ldots, 0.85, 0.9\}} \max_{k \in \{0, 0.01, \ldots, \lfloor \frac{K_u(\tau)}{0.01} \rfloor \times 0.01\}} \frac{|\hat{\pi}_U^{80}(\tau, k) - \tilde{\pi}_U^{80}(\tau, k)|}{\omega(\tau)}, \tag{C.9}$$

which we denote by $\tilde{r}_{1-\alpha}^{80}$. Similarly, we compute the $1-\alpha$ quantile across bootstrap samples of

$$\max_{\tau \in \{0.1, 0,15, \ldots, 0.85, 0.9\}} \max_{k \in \{0, 0.01, \ldots, \lfloor \frac{K_u(\tau)}{0.01} \rfloor \times 0.01\}} \frac{|\hat{\pi}_L^{90}(\tau, k) - \tilde{\pi}_L^{90}(\tau, k)|}{\omega(\tau)}, \tag{C.10}$$

which we denote by $\tilde{r}_{1-\alpha}^{90}$. We then let $k_{I,L}(\tau)$ be the smallest value of k in the grid such that

$$\hat{\pi}_U^{80}(\tau, k) + \tilde{r}_{1-\alpha}^{80} \omega(\tau) \geq \hat{\pi}_L^{90}(\tau, k) - \tilde{r}_{1-\alpha}^{90} \omega(\tau). \tag{C.11}$$

The lower confidence band is then

$$\hat{\kappa}_L(\tau) \equiv \left(\hat{\pi}_U^{80}\left(\tau, k_{I,L}(\tau)\right) + \tilde{r}_{1-\alpha}^{80} \omega(\tau) + \hat{\pi}_L^{90}\left(\tau, k_{I,L}(\tau)\right) - \tilde{r}_{1-\alpha}^{90} \omega(\tau)\right)/2.$$

Analogously, we let $k_{I,U}(\tau)$ be the smallest value of $k \in \{0, 0.001, \ldots, \lfloor \frac{K_u(\tau)}{0.001} \rfloor \times 0.001\}$ such that

$$\hat{\pi}_U^{80}(\tau, k) - \tilde{r}_{1-\alpha}^{80} \omega(\tau) \geq \hat{\pi}_L^{90}(\tau, k) + \tilde{r}_{1-\alpha}^{90} \omega(\tau), \tag{C.12}$$

and get the upper confidence band

$$\hat{\kappa}_U(\tau) \equiv \left(\hat{\pi}_U^{80}\left(\tau, k_{I,U}(\tau)\right) - \tilde{r}_{1-\alpha}^{80} \omega(\tau) + \hat{\pi}_L^{90}\left(\tau, k_{I,U}(\tau)\right) + \tilde{r}_{1-\alpha}^{90} \omega(\tau)\right)/2.$$

Figure 9 is a graph of $\hat{\kappa}_L(\tau)$, $\hat{\kappa}_U(\tau)$, and $\hat{\kappa}(\tau)$. Our bootstraps were conducted over a coarser grid than the one used to obtain point estimates so as to save on computational cost.

FIGURE 10. Here λ is set to different levels of education and all other coordinates are set equal to the sample mean of $\{X_i\}_{i=1}^n$. The procedure is otherwise identical to the one employed in the construction of Figure 7, with the exception that a quantile specific critical value is employed.

APPENDIX D: DERIVATIONS OF SECTION 5.2

This appendix provides a justification for the derivations in Section 5.2, in particular, derivations of the representation derived in equation (62). Toward this end, observe first that by Bayes' rule,

$$
\begin{aligned}
F_{y|1,x}(c) &= \frac{P(D=1|X=x, Y \le c) \times F_{y|x}(c)}{p(x)} \\
&= \frac{P(D=1|X=x, F_{y|x}(Y) \le F_{y|x}(c)) \times F_{y|x}(c)}{p(x)},
\end{aligned} \tag{D.1}
$$

where the second equality follows from $F_{y|x}$ being strictly increasing. Evaluating (D.1) at $c = q(\tau|x)$, employing the definition of $p_L(x,\tau)$, and noting that $F_{y|x}(q(\tau|x)) = \tau$ yields

$$
F_{y|1,x}\big(q(\tau|x)\big) = \frac{p_L(\tau,x) \times \tau}{p(x)}. \tag{D.2}
$$

Moreover, by identical arguments, but working instead with the definition of $p_U(\tau, x)$, we derive

$$
\begin{aligned}
1 - F_{y|1,x}\big(q(\tau|x)\big) &= \frac{P(D=1|Y > q(\tau|x), X=x) \times (1 - F_{y|1,x}(q(\tau|x)))}{p(x)} \\
&= \frac{p_U(\tau,x) \times (1-\tau)}{p(x)}.
\end{aligned} \tag{D.3}
$$

Finally, we note that the same manipulations applied to $F_{y|0,x}$ instead of $F_{y|1,x}$ enable us to obtain

$$
\begin{aligned}
F_{y|0,x}\big(q(\tau|x)\big) &= \frac{(1 - p_L(\tau,x)) \times \tau}{1 - p(x)}, \\
1 - F_{y|0,x}\big(q(\tau|x)\big) &= \frac{(1 - p_U(\tau,x)) \times (1-\tau)}{1 - p(x)}.
\end{aligned} \tag{D.4}
$$

Hence, we can obtain by direct algebra from the results (D.1) and (D.4) that we must have

$$
\big|F_{y|1,x}\big(q(\tau|x)\big) - F_{y|0,x}\big(q(\tau|x)\big)\big| = \frac{|p(x) - p_L(x,\tau)| \times \tau}{p(x)(1-p(x))}. \tag{D.5}
$$

Analogously, exploiting (D.1) and (D.4) once again, we can also obtain

$$
\begin{aligned}
&\big|F_{y|1,x}\big(q(\tau|x)\big) - F_{y|0,x}\big(q(\tau|x)\big)\big| \\
&\quad = \big|\big(1 - F_{y|1,x}\big(q(\tau|x)\big)\big) - \big(1 - F_{y|0,x}\big(q(\tau|x)\big)\big)\big| \\
&\quad = \frac{|p(x) - p_U(x,\tau)| \times (1-\tau)}{p(x)(1-p(x))}.
\end{aligned} \tag{D.6}
$$

Equation (62) then follows from taking the square root of the product of (D.5) and (D.6).

Appendix E: Proof of Results

Proof of Lemma 3.1. For any $\theta \in \mathcal{C}(\tau, k)$, the first order condition of the norm minimization problem yields $\beta(\tau) = (E_S[X_i X_i'])^{-1} E_S[X_i \theta(X_i)]$. The lemma then follows from Corollary 2.1. □

PROOF OF COROLLARY 3.1. Since $\mathcal{P}(\tau, k)$ is convex by Lemma 3.1, it follows that the identified set for $\lambda'\beta(\tau)$ is a convex set in $\mathbf{R}$ and, hence, is an interval. The fact that $\pi_L(\tau, k)$ and $\pi_U(\tau, k)$ are the endpoints of such an interval follows directly from Lemma 3.1. □

LEMMA E.1. *Let Assumption* 2.1 *hold, and let* W_i *be independent of* (Y_i, X_i, D_i) *with* $E[W_i] = 1$ *and* $E[W_i^2] < \infty$ *and positive almost surely. If* $\{Y_i D_i, X_i, D_i, W_i\}$ *is an i.i.d. sample, then the following class is Donsker*:

$$\mathcal{M} \equiv \big\{m_c : m_c(y, x, d, w) \equiv w1\{y \le c, d = 1, x = x_0\} - P(Y_i \le c, D_i = 1, X_i = x_0),\; c \in \mathbf{R}\big\}.$$

PROOF. For any $1 > \varepsilon > 0$, there is an increasing sequence $-\infty = y_0 \le \cdots \le y_{\lceil 8/\varepsilon\rceil} = +\infty$ such that for any $1 \le j \le \lceil \frac{8}{\varepsilon} \rceil$, we have $F_{y|1,x}(y_j) - F_{y|1,x}(y_{j-1}) < \varepsilon/4$. Next, define the functions

$$l_j(y, x, d, w) \equiv w1\{y \le y_{j-1}, d = 1, x = x_0\} - P(Y_i \le y_j, D_i = 1, X_i = x_0), \tag{E.1}$$

$$u_j(y, x, d, w) \equiv w1\{y \le y_j, d = 1, x = x_0\} - P(Y_i \le y_{j-1}, D_i = 1, X_i = x_0) \tag{E.2}$$

and notice that the brackets $\{[l_j, u_j]\}_{j=1}^{\lceil 8/\varepsilon\rceil}$ form a partition of the class $\mathcal{M}_c$ (since $w \in \mathbf{R}_+$). In addition, note that

$$\begin{aligned} &E\big[\big(u_j(Y_i, X_i, D_i, W_i) - l_j(Y_i, X_i, D_i, W_i)\big)^2\big] \\ &\quad\le 2E\big[W_i^2 1\{y_{j-1} \le Y_i \le y_j, D_i = 1, X_i = x_0\}\big] \\ &\qquad + 2P^2(y_{j-1} \le Y_i \le y_j, D_i = 1, X_i = x_0) \\ &\quad\le 4E\big[W_i^2\big] \times \big(F_{y|1,x}(y_j) - F_{y|1,x}(y_{j-1})\big), \end{aligned} \tag{E.3}$$

and, hence, each bracket has norm bounded by $\sqrt{E[W_i^2]\varepsilon}$. Therefore, $N_{[]}(\varepsilon, \mathcal{M}, \|\cdot\|_{L^2}) \le 16E[W_i^2]/\varepsilon^2$, and the lemma follows by Theorem 2.5.6 in van der Vaart and Wellner (1996). □

LEMMA E.2. *Let Assumption* 2.1 *hold, and let* W_i *be independent of* (Y_i, X_i, D_i) *with* $E[W_i] = 1$, $E[W_i^2] < \infty$ *and positive almost surely. Also let* $\mathcal{S}_\varepsilon \equiv \{(\tau, b) \in [0, 1]^2 : b\{1 - p(x)\} + \varepsilon \le \tau \le p(x) + b\{1 - p(x)\} - \varepsilon\ \forall x \in \mathcal{X}\}$ *for some* ε *satisfying* $0 < 2\varepsilon < \inf_{x\in\mathcal{X}} p(x)$ *and denote the minimizers*

$$s_0(\tau, b, x) = \arg\min_{c\in\mathbf{R}} Q_x(c|\tau, b), \qquad \hat{s}_0(\tau, b, x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\tau, b). \tag{E.4}$$

Then $s_0(\tau, b, x)$ *is bounded in* $(\tau, b, x) \in \mathcal{S}_\varepsilon \times \mathcal{X}$ *and if* $\{Y_i D_i, X_i, D_i, W_i\}$ *is i.i.d., then for some* $M > 0$,

$$P\Big(\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \big|\hat{s}_0(\tau, b, x)\big| > M\Big) = o(1).$$

Proof. First note that Assumption 2.1(ii) implies $s_0(\tau, b, x)$ is uniquely defined, while $\hat{s}_0(\tau, b, x)$ may be one of multiple minimizers. By Assumption 2.1(ii) and the definition of $\mathcal{S}_\varepsilon$, it follows that the equality

$$P\big(Y_i \leq s_0(\tau, b, x), D_i = 1 | X_i = x\big) = \tau - bP(D_i = 0 | X_i = x) \tag{E.5}$$

implicitly defines $s_0(\tau, b, x)$. Let $\bar{s}(x)$ and $\underline{s}(x)$ be the unique numbers satisfying $F_{y|1,x}(\bar{s}(x)) \times p(x) = p(x) - \varepsilon$ and $F_{y|1,x}(\underline{s}(x)) \times p(x) = \varepsilon$. By result (E.5) and the definition of $\mathcal{S}_\varepsilon$, we then obtain that for all $x \in \mathcal{X}$,

$$-\infty < \underline{s}(x) \leq \inf_{(\tau,b)\in\mathcal{S}_\varepsilon} s_0(\tau, b, x) \leq \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} s_0(\tau, b, x) \leq \bar{s}(x) < +\infty. \tag{E.6}$$

Hence, we conclude that $\sup_{(\tau,b)\in\mathcal{S}_\varepsilon} |s_0(\tau, b, x)| = O(1)$ and the first claim follows by $\mathcal{X}$ being finite.

To establish the second claim of the lemma, we define the functions

$$R_x(\tau, b) \equiv bP(D_i = 0, X_i = x) - \tau P(X_i = x), \tag{E.7}$$

$$R_{x,n}(\tau, b) \equiv \frac{1}{n}\sum_{i=1}^{n} W_i\big(b1\{D_i = 0, X_i = x\} - \tau 1\{X_i = x\}\big) \tag{E.8}$$

as well as the maximizers and the minimizers of $R_{x,n}(\tau, b)$ on the set $\mathcal{S}_\varepsilon$, which we denote by

$$\begin{aligned}\big(\underline{\tau}_n(x), \underline{b}_n(x)\big) &\in \arg\max_{(\tau,b)\in\mathcal{S}_\varepsilon} R_{x,n}(\tau, b),\\ \big(\bar{\tau}_n(x), \bar{b}_n(x)\big) &\in \arg\min_{(\tau,b)\in\mathcal{S}_\varepsilon} R_{x,n}(\tau, b).\end{aligned} \tag{E.9}$$

Also denote the set of maximizers and minimizers of $\tilde{Q}_{x,n}(c|\tau, b)$ at these particular choices of (τ, b) by

$$\underline{S}_n(x) \equiv \Big\{\underline{s}_n(x) \in \mathbf{R} : \underline{s}_n(x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}\big(c|\underline{\tau}_n(x), \underline{b}_n(x)\big)\Big\}, \tag{E.10}$$

$$\bar{S}_n(x) \equiv \Big\{\bar{s}_n(x) \in \mathbf{R} : \bar{s}_n(x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}\big(c|\bar{\tau}_n(x), \bar{b}_n(x)\big)\Big\}. \tag{E.11}$$

From the definition of $\tilde{Q}_{x,n}(c|\tau, b)$, we then obtain from (E.9), (E.10), and (E.11) that for all $x \in \mathcal{X}$,

$$\inf_{\underline{s}_n(x)\in\underline{S}_n(x)} \underline{s}_n(x) \leq \inf_{(\tau,b)\in\mathcal{S}_\varepsilon} \hat{s}_0(\tau, b, x) \leq \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \hat{s}_0(\tau, b, x) \leq \sup_{\bar{s}_n(x)\in\bar{S}_n(x)} \bar{s}_n(x). \tag{E.12}$$

We establish the second claim of the lemma by exploiting (E.12) and showing that for some $0 < M < \infty$,

$$P\Big(\inf_{\underline{s}_n(x)\in\underline{S}_n(x)} \underline{s}_n(x) < -M\Big) = o(1), \qquad P\Big(\sup_{\bar{s}_n(x)\in\bar{S}_n(x)} \bar{s}_n(x) > M\Big) = o(1). \tag{E.13}$$

To prove that $\inf_{\underline{s}_n(x)\in\underline{S}_n(x)} \underline{s}_n(x)$ is larger than $-M$ with probability tending to 1, note that

$$\begin{aligned}&\big|R_{x,n}\big(\underline{\tau}_n(x),\underline{b}_n(x)\big)+\varepsilon P(X_i=x)\big|\\ &\quad=\Big|R_{x,n}\big(\underline{\tau}_n(x),\underline{b}_n(x)\big)-\max_{(\tau,b)\in\mathcal{S}_\varepsilon} R_x(\tau,b)\Big|=o_p(1),\end{aligned}\tag{E.14}$$

where the second equality follows from the theorem of the maximum and the continuous mapping theorem. Therefore, using the equality $a^2-b^2=(a-b)(a+b)$, result (E.14), and Lemma E.1, it follows that

$$\sup_{c\in\mathbf{R}}\big|\tilde{Q}_{x,n}\big(c|\underline{\tau}_n(x),\underline{b}_n(x)\big)-\big(F_{y|1,x}(c)p(x)-\varepsilon\big)^2P^2(X_i=x)\big|=o_p(1).\tag{E.15}$$

Fix $\delta>0$ and note that since $F_{y|1,x}(\underline{s}(x))p(x)=\varepsilon$ and $\varepsilon/p(x)<1$, Assumption 2.1(ii) implies that

$$\eta\equiv\inf_{|c-\underline{s}(x)|>\delta}\big(F_{y|1,x}(c)p(x)-\varepsilon\big)^2>0.\tag{E.16}$$

Therefore, it follows from direct manipulations, and the definition of $\underline{S}_n(x)$ in (E.10) and of $\underline{s}(x)$ that

$$\begin{aligned}&P\Big(\Big|\inf_{\underline{s}_n(x)\in\underline{S}_n(x)}\underline{s}_n(x)-\underline{s}(x)\Big|>\delta\Big)\\ &\quad\le P\Big(\inf_{|c-\underline{s}(x)|>\delta}\tilde{Q}_{x,n}\big(c|\underline{\tau}_n(x),\underline{b}_n(x)\big)\le\tilde{Q}_{x,n}\big(\underline{s}(x)|\underline{\tau}_n(x),\underline{b}_n(x)\big)\Big)\\ &\quad\le P\Big(\eta\le\sup_{c\in\mathbf{R}}2\big|\tilde{Q}_{x,n}\big(c|\underline{\tau}_n(x),\underline{b}_n(x)\big)-\big(F_{y|1,x}(c)p(x)-\varepsilon\big)^2P^2(X_i=x)\big|\Big).\end{aligned}$$

We hence conclude from (E.15) that $\inf_{\underline{s}_n(x)\in\underline{S}_n(x)}\underline{s}_n(x)\xrightarrow{p}\underline{s}(x)$, which together with (E.6) implies that $\inf_{\underline{s}_n(x)\in\underline{S}_n(x)}\underline{s}_n(x)$ is larger than $-M$ with probability tending to 1 for some $M>0$. By similar arguments, it can be shown that $\sup_{\bar{s}_n(x)\in\bar{S}_n(x)}\bar{s}_n(x)\xrightarrow{p}\bar{s}(x)$, which together with (E.6) establishes (E.13). The second claim of the lemma then follows from (E.12), (E.13), and $\mathcal{X}$ being finite. □

Lemma E.3. *Let Assumption* 2.1 *hold, and let W_i be independent of (Y_i,X_i,D_i) with $E[W_i]=1$, $E[W_i^2]<\infty$ and positive almost surely. Also let $\mathcal{S}_\varepsilon\equiv\{(\tau,b)\in[0,1]^2: b\{1-p(x)\}+\varepsilon\le\tau\le p(x)+b\{1-p(x)\}-\varepsilon\ \forall x\in\mathcal{X}\}$ for some ε satisfying $0<2\varepsilon<\inf_{x\in\mathcal{X}}p(x)$ and denote the minimizers*

$$s_0(\tau,b,x)=\arg\min_{c\in\mathbf{R}}Q_x(c|\tau,b),\qquad \hat{s}_0(\tau,b,x)\in\arg\min_{c\in\mathbf{R}}\tilde{Q}_{x,n}(c|\tau,b).\tag{E.17}$$

If $\{Y_iD_i,X_i,D_i,W_i\}$ is an i.i.d. sample, then $\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}|\hat{s}_0(\tau,b,x)-s_0(\tau,b,x)|=o_p(1)$.

Proof. First define the criterion functions $M : L^\infty(\mathcal{S}_\varepsilon \times \mathcal{X}) \to \mathbf{R}$ and $M_n : L^\infty(\mathcal{S}_\varepsilon \times \mathcal{X}) \to \mathbf{R}$ by

$$\begin{aligned} M(\theta) &\equiv \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} Q_x\big(\theta(\tau,b,x)|\tau,b\big), \\ M_n(\theta) &\equiv \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \tilde{Q}_{x,n}\big(\theta(\tau,b,x)|\tau,b\big). \end{aligned} \tag{E.18}$$

For notational convenience, let $s_0 \equiv s_0(\cdot,\cdot,\cdot)$ and $\hat{s}_0 \equiv \hat{s}_0(\cdot,\cdot,\cdot)$. By Lemma E.2, $s_0 \in L^\infty(\mathcal{S}_\varepsilon \times \mathcal{X})$, while with probability tending to 1, $\hat{s}_0 \in L^\infty(\mathcal{S}_\varepsilon \times \mathcal{X})$. Hence, (E.17) implies that with probability tending to 1,

$$\hat{s}_0 \in \arg\min_{\theta\in L^\infty(\mathcal{S}_\varepsilon\times\mathcal{X})} M_n(\theta), \qquad s_0 = \arg\min_{\theta\in L^\infty(\mathcal{S}_\varepsilon\times\mathcal{X})} M(\theta). \tag{E.19}$$

By Assumption 2.1(ii) and (E.5), $Q_x(\cdot|\tau,b)$ is strictly convex in a neighborhood of $s_0(\tau,b,x)$. Furthermore, since by (E.5) and the implicit function theorem, $s_0(\tau,b,x)$ is continuous in $(\tau,b)\in\mathcal{S}_\varepsilon$ for every $x\in\mathcal{X}$, then

$$\begin{aligned} &\inf_{\|\theta-s_0\|_\infty\geq\delta} M(\theta) \\ &\quad\geq \inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}_\varepsilon} \inf_{|c-s_0(\tau,b,x)|\geq\delta} Q_x(c|\tau,b) \\ &\quad= \inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}_\varepsilon} \min\big\{Q_x\big(s_0(\tau,b,x)-\delta|\tau,b\big), Q_x\big(s_0(\tau,b,x)+\delta|\tau,b\big)\big\} \\ &\quad> 0, \end{aligned} \tag{E.20}$$

where the final inequality follows by compactness of $\mathcal{S}_\varepsilon$, which together with continuity of $s_0(\tau,b,x)$, implies the inner infimum is attained, while the outer infimum is trivially attained due to $\mathcal{X}$ being finite. Since (E.20) holds for any $\delta>0$, s_0 is a well separated minimum of $M(\theta)$ in $L^\infty(\mathcal{S}_\varepsilon\times\mathcal{X})$. Next define

$$G_{x,i}(c) \equiv W_i 1\{Y_i \leq c, D_i = 1, X_i = x\} \tag{E.21}$$

and observe that compactness of $\mathcal{S}_\varepsilon$, a regular law of large numbers, Lemma E.1, and finiteness of $\mathcal{X}$ yield

$$\begin{aligned} &\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \sup_{c\in\mathbf{R}} \left| \frac{1}{n}\sum_{i=1}^n G_{x,i}(c) + R_{x,n}(\tau,b) - E\big[G_{x,i}(c)\big] - R_x(\tau,b) \right| \\ &\quad\leq \sup_{x\in\mathcal{X}} \sup_{c\in\mathbf{R}} \left| \frac{1}{n}\sum_{i=1}^n G_{x,i}(c) - E\big[G_{x,i}(c)\big] \right| \\ &\qquad + \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \big|R_{x,n}(\tau,b) - R_x(\tau,b)\big| \\ &\quad= o_p(1), \end{aligned} \tag{E.22}$$

where $R_x(\tau, b)$ and $R_{x,n}(\tau, b)$ are as in (E.7) and (E.8), respectively. Hence, using (E.22), the equality $a^2 - b^2 = (a-b)(a+b)$ and $Q_x(c|\tau, b)$ uniformly bounded in $(c, \tau, b) \in \mathbf{R} \times \mathcal{S}_\varepsilon$ due to the compactness of $\mathcal{S}_\varepsilon$, we obtain

$$\begin{aligned}
&\sup_{\theta \in L^\infty(\mathcal{S}_\varepsilon \times \mathcal{X})} \big| M_n(\theta) - M(\theta) \big| \\
&\quad \le \sup_{\theta \in L^\infty(\mathcal{S}_\varepsilon \times \mathcal{X})} \sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}_\varepsilon} \big| \tilde{Q}_{x,n}\big(\theta(\tau, b, x)|\tau, b\big) - Q_x\big(\theta(\tau, b, x)|\tau, b\big) \big| \\
&\quad \le \sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}_\varepsilon} \sup_{c \in \mathbf{R}} \big| \tilde{Q}_{x,n}(c|\tau, b) - Q_x(c|\tau, b) \big| \\
&\quad = o_p(1).
\end{aligned} \tag{E.23}$$

The claim of the lemma then follows from results (E.19), (E.20), and (E.23) together with Corollary 3.2.3 in van der Vaart and Wellner (1996). □

Lemma E.4. *Let Assumptions* 2.1 *and* 4.1 *hold, and let* W_i *be independent of* (Y_i, X_i, D_i) *with* $E[W_i] = 1$, $E[W_i^2] < \infty$, *and positive a.s. Also let* $\mathcal{S}_\varepsilon \equiv \{(\tau, b) \in [0,1]^2 : b\{1 - p(x)\} + \varepsilon \le \tau \le p(x) + b\{1 - p(x)\} - \varepsilon \ \forall x \in \mathcal{X}\}$ *for some* ε *satisfying* $0 < 2\varepsilon < \inf_{x \in \mathcal{X}} p(x)$ *and denote the minimizers*

$$s_0(\tau, b, x) = \arg\min_{c \in \mathbf{R}} Q_x(c|\tau, b), \qquad \hat{s}_0(\tau, b, x) \in \arg\min_{c \in \mathbf{R}} \tilde{Q}_{x,n}(c|\tau, b). \tag{E.24}$$

For $G_{x,i}(c) \equiv W_i 1\{Y_i \le c, D_i = 1, X_i = x\}$ *and* $R_{x,n}(\tau, b)$ *as defined in* (E.8), *denote the criterion function*

$$\begin{aligned}
\tilde{Q}^s_{x,n}(c|\tau, b) \equiv \Bigg(&\frac{1}{n} \sum_{i=1}^n \big\{ E\big[G_{x,i}(c) - G_{x,i}\big(s_0(\tau, b, x)\big)\big] + G_{x,i}\big(s_0(\tau, b, x)\big) \big\} \\
&+ R_{x,n}(\tau, b) \Bigg)^2.
\end{aligned} \tag{E.25}$$

If $\{Y_i D_i, X_i, D_i, W_i\}$ *is an i.i.d. sample, it then follows that*

$$\sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}_\varepsilon} \left| \frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau, b, x)|\tau, b)}{dc} \right| = o_p\big(n^{-1/2}\big). \tag{E.26}$$

Proof. We first introduce the criterion function $M^s_n : L^\infty(\mathcal{S}_\varepsilon \times \mathcal{X}) \to \mathbf{R}$ to be given by

$$M^s_n(\theta) \equiv \sup_{x \in \mathcal{X}} \sup_{(\tau,b) \in \mathcal{S}_\varepsilon} \tilde{Q}^s_{x,n}\big(\theta(\tau, b, x)|\tau, b\big). \tag{E.27}$$

We aim to characterize and establish the consistency of an approximate minimizer of $M^s_n(\theta)$ on $L^\infty(\mathcal{S}_\varepsilon \times \mathcal{X})$. Observe that by Lemma E.1, compactness of $\mathcal{S}_\varepsilon$, finiteness of $\mathcal{X}$, and the law of large numbers,

$$\begin{aligned}
&\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\left|\frac{1}{n}\sum_{i=1}^n\{G_{x,i}\big(s_0(\tau,b,x)\big)-E\big[G_{x,i}\big(s_0(\tau,b,x)\big)\big]\}\right.\\
&\qquad\left.+R_{x,n}(\tau,b)-R_x(\tau,b)\right|\\
&\quad\le\sup_{x\in\mathcal{X}}\sup_{c\in\mathbf{R}}\left|\frac{1}{n}\sum_{i=1}^n\{G_{x,i}(c)-E\big[G_{x,i}(c)\big]\}\right| \qquad\qquad \text{(E.28)}\\
&\qquad+\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\big|R_{x,n}(\tau,b)-R_x(\tau,b)\big|\\
&\quad=o_p(1),
\end{aligned}$$

where $R_x(\tau,b)$ is as in (E.7). Hence, by definition of $\mathcal{S}_\varepsilon$ and $R_x(\tau,b)$, with probability tending to 1,

$$\begin{aligned}
\frac{\varepsilon}{2}P(X_i=x)&\le\frac{1}{n}\sum_{i=1}^n\{E\big[G_{x,i}\big(s_0(\tau,b,x)\big)\big]-G_{x,i}\big(s_0(\tau,b,x)\big)\}-R_{x,n}(\tau,b)\\
&\le\left(p(x)-\frac{\varepsilon}{2}\right)P(X_i=x)\quad\forall(\tau,b,x)\in\mathcal{S}_\varepsilon\times\mathcal{X}.
\end{aligned} \tag{E.29}$$

By Assumption 2.1(ii), whenever (E.29) holds, we may implicitly define $\hat{s}_0^s(\tau,b,x)$ by the equality

$$\begin{aligned}
&P\big(Y_i\le\hat{s}_0^s(\tau,b,x),D_i=1,X_i=x\big)\\
&\quad=\frac{1}{n}\sum_{i=1}^n\{E\big[G_{x,i}\big(s_0(\tau,b,x)\big)\big]-G_{x,i}\big(s_0(\tau,b,x)\big)\}-R_{x,n}(\tau,b)
\end{aligned} \tag{E.30}$$

for all $(\tau,b,x)\in\mathcal{S}_\varepsilon\times\mathcal{X}$ and set $\hat{s}_0^s(\tau,b,x)=0$ for all $(\tau,b,x)\in\mathcal{S}_\varepsilon\times\mathcal{X}$ whenever (E.29) does not hold. Thus,

$$\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\left|\tilde{Q}_{x,n}^s\big(\hat{s}_0^s(\tau,b,x)|\tau,b\big)-\inf_{c\in\mathbf{R}}\tilde{Q}_{x,n}^s(c|\tau,b)\right|=o_p\big(n^{-1}\big). \tag{E.31}$$

Let $\hat{s}_0^s\equiv\hat{s}_0^s(\cdot,\cdot,\cdot)$ and note that by construction $\hat{s}_0^s\in L^\infty(\mathcal{S}_\varepsilon\times\mathcal{X})$. From (E.31), we then obtain that

$$\begin{aligned}
M_n^s\big(\hat{s}_0^s\big)&\le\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\inf_{c\in\mathbf{R}}\tilde{Q}_{x,n}^s(c|\tau,b)+o_p\big(n^{-1}\big)\\
&\le\inf_{\theta\in L^\infty(\mathcal{S}_\varepsilon\times\mathcal{X})}M_n^s(\theta)+o_p\big(n^{-1}\big).
\end{aligned} \tag{E.32}$$

To establish $\|\hat{s}_0^s - s_0\|_\infty = o_p(1)$, let $M(\theta)$ be as in (E.18) and notice that arguing as in (E.23) together with result (E.28) and Lemma E.1 implies that

$$\begin{aligned}
&\sup_{\theta\in L^\infty(\mathcal{S}_\varepsilon\times\mathcal{X})} \left|M_n^s(\theta) - M(\theta)\right| \\
&\quad\le \sup_{\theta\in L^\infty(\mathcal{S}_\varepsilon\times\mathcal{X})} \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \left|\tilde{Q}_{x,n}^s\big(\theta(\tau,b,x)|\tau,b\big) - Q_x\big(\theta(\tau,b,x)|\tau,b\big)\right| \\
&\quad\le \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \sup_{c\in\mathbf{R}} \left|\tilde{Q}_{x,n}^s(c|\tau,b) - Q_x(c|\tau,b)\right| \\
&\quad= o_p(1).
\end{aligned}\tag{E.33}$$

Hence, by (E.20), (E.32), (E.33), and Corollary 3.2.3 in van der Vaart and Wellner (1996), we obtain

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \left|\hat{s}_0^s(\tau,b,x) - s_0(\tau,b,x)\right| = o_p(1).\tag{E.34}$$

Next, define the random mapping $\Delta_n : L^\infty(\mathcal{S}_\varepsilon\times\mathcal{X}) \to L^\infty(\mathcal{S}_\varepsilon\times\mathcal{X})$ to be given by

$$\begin{aligned}
\Delta_n(\theta)(\tau,b,x) \equiv \frac{1}{n}\sum_{i=1}^n \Big\{&\Big(G_{x,i}\big(\theta(\tau,b,x)\big) - E\big[G_{x,i}\big(\theta(\tau,b,x)\big)\big]\Big) \\
&- \Big(G_{x,i}\big(s_0(\tau,b,x)\big) - E\big[G_{x,i}\big(s_0(\tau,b,x)\big)\big]\Big)\Big\},
\end{aligned}\tag{E.35}$$

and observe that Lemma E.1 and finiteness of $\mathcal{X}$ imply that $\|\Delta_n(\bar{s})\|_\infty = o_p(n^{-1/2})$ for any $\bar{s} \in L^\infty(\mathcal{S}_\varepsilon\times\mathcal{X})$ such that $\|\bar{s} - s_0\|_\infty = o_p(1)$. Since $\tilde{Q}_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b) \le \tilde{Q}_{x,n}(s_0(\tau,b,x)|\tau,b)$ for all $(\tau,b,x)\in\mathcal{S}_\varepsilon\times\mathcal{X}$, and by Lemma E.1 and finiteness of $\mathcal{X}$, $\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \tilde{Q}_{x,n}(s_0(\tau,b,x)|\tau,b) = O_p(n^{-1})$, we conclude that

$$\begin{aligned}
&\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \left\{\tilde{Q}_{x,n}^s\big(\hat{s}_0(\tau,b,x)|\tau,b\big) - \tilde{Q}_{x,n}^s\big(\hat{s}_0^s(\tau,b,x)|\tau,b\big)\right\} \\
&\quad\le \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \left\{\tilde{Q}_{x,n}\big(\hat{s}_0(\tau,b,x)|\tau,b\big) - \tilde{Q}_{x,n}^s\big(\hat{s}_0^s(\tau,b,x)|\tau,b\big)\right\} \\
&\qquad+ \left\|\Delta_n^2(\hat{s}_0)\right\|_\infty + 2\left\|\Delta_n(\hat{s}_0)\right\|_\infty \times M_n^{1/2}(\hat{s}_0) \\
&\quad\le \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \left\{\tilde{Q}_{x,n}\big(\hat{s}_0^s(\tau,b,x)|\tau,b\big) - \tilde{Q}_{x,n}^s\big(\hat{s}_0^s(\tau,b,x)|\tau,b\big)\right\} + o_p\big(n^{-1}\big),
\end{aligned}\tag{E.36}$$

where $M_n(\theta)$ is as in (E.18). Furthermore, since by (E.32), we have $M_n^s(\hat{s}_0^s) \le M_n^s(s_0) + o_p(n^{-1})$, and by Lemma E.1 and finiteness of $\mathcal{X}$ we have $M_n^s(s_0) = O_p(n^{-1})$, similar arguments as in (E.36) imply that

$$\begin{aligned}
&\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \left\{\tilde{Q}_{x,n}\big(\hat{s}_0^s(\tau,b,x)|\tau,b\big) - \tilde{Q}_{x,n}^s\big(\hat{s}_0^s(\tau,b,x)|\tau,b\big)\right\} \\
&\quad\le \left\|\Delta_n\big(\hat{s}_0^s\big)\right\|_\infty^2 + 2\left\|\Delta_n\big(\hat{s}_0^s\big)\right\|_\infty \times \left[M_n^s\big(\hat{s}_0^s\big)\right]^{1/2} = o_p\big(n^{-1}\big).
\end{aligned}\tag{E.37}$$

Therefore, by combining the results in (E.31), (E.36), and (E.37), we are able to conclude that

$$\begin{aligned}&\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\Big\{\tilde{Q}^s_{x,n}\big(\hat{s}_0(\tau,b,x)|\tau,b\big)-\inf_{c\in\mathbf{R}}\tilde{Q}^s_{x,n}(c|\tau,b)\Big\}\\&\leq\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\big\{\tilde{Q}^s_{x,n}\big(\hat{s}_0(\tau,b,x)|\tau,b\big)-\tilde{Q}^s_{x,n}\big(\hat{s}^s_0(\tau,b,x)|\tau,b\big)\big\}+o_p\big(n^{-1}\big)\\&\leq o_p\big(n^{-1}\big).\end{aligned}\tag{E.38}$$

Let $\varepsilon_n\searrow 0$ be such that $\varepsilon_n=o_p(n^{-1/2})$ and in addition satisfies

$$\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\Big|\tilde{Q}^s_{x,n}\big(\hat{s}_0(\tau,b,x)|\tau,b\big)-\inf_{c\in\mathbf{R}}\tilde{Q}^s_{x,n}(c|\tau,b)\Big|=o_p\big(\varepsilon_n^2\big),\tag{E.39}$$

which is possible by (E.38). A Taylor expansion at each $(\tau,b,x)\in\mathcal{S}_\varepsilon\times\mathcal{X}$ then implies

$$\begin{aligned}0&\leq\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\big\{\tilde{Q}^s_{x,n}\big(\hat{s}_0(\tau,b,x)+\varepsilon_n|\tau,b\big)-\tilde{Q}^s_{x,n}\big(\hat{s}_0(\tau,b,x)|\tau,b\big)\big\}\\&\quad+o_p\big(\varepsilon_n^2\big)\\&=\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in S}\left\{\varepsilon_n\times\frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc}+\frac{\varepsilon_n^2}{2}\times\frac{d^2\tilde{Q}^s_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2}\right\}\\&\quad+o_p\big(\varepsilon_n^2\big),\end{aligned}\tag{E.40}$$

where $\bar{s}(\tau,b,x)$ is a convex combination of $\hat{s}_0(\tau,b,x)$ and $\hat{s}_0(\tau,b,x)+\varepsilon_n$. Since Lemma E.3 and $\varepsilon_n\searrow 0$ imply that $\|\bar{s}-s_0\|_\infty=o_p(1)$, the mean value theorem, $f_{y|1,x}(c)$ being uniformly bounded, and (E.23) yield

$$\begin{aligned}&\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\bigg|\frac{1}{n}\sum_{i=1}^n\big\{E\big[G_{x,i}\big(\bar{s}(\tau,b,x)\big)-G_{x,i}\big(s_0(\tau,b,x)\big)\big]+G_{x,i}\big(s_0(\tau,b,x)\big)\big\}\\&\qquad+R_{x,n}(\tau,b)\bigg|\\&\leq\sup_{c\in\mathbf{R}}f_{y|1,x}(c)p(x)P(X_i=x)\times\|\bar{s}-s_0\|_\infty+M_n^{1/2}(s_0)=o_p(1).\end{aligned}\tag{E.41}$$

Therefore, exploiting (E.41), $f'_{y|1,x}(c)$ being uniformly bounded, and by direct calculation, we conclude that

$$\begin{aligned}&\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\left|\frac{d^2\tilde{Q}^s_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2}-2f^2_{y|1,x}\big(\bar{s}(\tau,b,x)\big)p^2(x)P^2(X_i=x)\right|\\&\leq\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\big|f'_{y|1,x}\big(\bar{s}(\tau,b,x)\big)p(x)P(X_i=x)\big|\times o_p(1)=o_p(1).\end{aligned}\tag{E.42}$$

Thus, combining results (E.40) together with (E.42) and $f_{y|1,x}(c)$ uniformly bounded, we conclude

$$0 \le \varepsilon_n \times \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} + O_p\big(\varepsilon_n^2\big). \tag{E.43}$$

In a similar fashion, we note that by exploiting (E.39) and proceeding as in (E.40)–(E.43), we obtain

$$\begin{aligned}
0 &\le \inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}_\varepsilon} \big\{\tilde{Q}^s_{x,n}\big(\hat{s}_0(\tau,b,x)-\varepsilon_n|\tau,b\big) - \tilde{Q}^s_{x,n}\big(\hat{s}_0(\tau,b,x)|\tau,b\big)\big\} + o_p\big(\varepsilon_n^2\big) \\
&\le \inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}_\varepsilon} \bigg\{-\varepsilon_n \times \frac{d\tilde{Q}_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} \\
&\quad + \frac{\varepsilon_n^2}{2} \times \frac{d^2\tilde{Q}^s_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2}\bigg\} + o_p\big(\varepsilon_n^2\big) \\
&\le -\varepsilon_n \times \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \frac{d\tilde{Q}_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} + O_p\big(\varepsilon_n^2\big).
\end{aligned} \tag{E.44}$$

Therefore, since $\varepsilon_n = o_p(n^{-1/2})$, we conclude from (E.43) and (E.44) that we must have

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} = O_p(\varepsilon_n) = o_p\big(n^{-1/2}\big). \tag{E.45}$$

By similar arguments, but reversing the sign of ε_n in (E.40) and (E.44), it possible to establish that

$$\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} -\frac{d\tilde{Q}^s_{x,n}(\hat{s}_0(\tau,b,x)|\tau,b)}{dc} = o_p\big(n^{-1/2}\big). \tag{E.46}$$

The claim of the lemma then follows from (E.45) and (E.46). □

LEMMA E.5. *Let Assumptions* 2.1 *and* 4.1 *hold, and let* W_i *be independent of* (Y_i, X_i, D_i) *with* $E[W_i] = 1$, $E[W_i^2] < \infty$, *and positive a.s. Also let* $\mathcal{S}_\varepsilon \equiv \{(\tau,b) \in [0,1]^2 : b\{1-p(x)\} + \varepsilon \le \tau \le p(x) + b\{1-p(x)\} - \varepsilon\ \forall x \in \mathcal{X}\}$ *for some* ε *satisfying* $0 < \varepsilon < \inf_{x\in\mathcal{X}} p(x)$ *and denote the minimizers*

$$s_0(\tau,b,x) = \arg\min_{c\in\mathbf{R}} Q_x(c|\tau,b), \qquad \hat{s}_0(\tau,b,x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\tau,b). \tag{E.47}$$

If $G_{x,i}(c)$ *is as in* (E.21), $\inf_{x\in\mathcal{X}} \inf_{(\tau,b)\in\mathcal{S}_\varepsilon} f_{y|1,x}(s_0(\tau,b,x))p(x) > 0$, *and* $\{Y_iD_i, X_i, D_i, W_i\}$ *is i.i.d., then*

$$\begin{aligned}
\sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\varepsilon} \bigg| &\big(\hat{s}_0(\tau,b,x) - s_0(\tau,b,x)\big) \\
&- \frac{1}{n}\sum_{i=1}^n \frac{G_{x,i}(s_0(\tau,b,x)) + W_i(b1\{D_i=0, X_i=x\} - \tau 1\{X_i=x\})}{P(X_i=x)p(x)f_{y|1,x}(s_0(\tau,b,x))}\bigg| = o_p\big(n^{-1/2}\big).
\end{aligned} \tag{E.48}$$

Proof. For $\tilde{Q}^s_{x,n}(c|\tau,b)$ as in (E.25), note that the mean value theorem and Lemma E.4 imply

$$\begin{aligned}\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\bigg|&\big(\hat{s}_0(\tau,b,x)-s_0(\tau,b,x)\big)\times\frac{d^2\tilde{Q}^s_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2}\\&+\frac{d\tilde{Q}^s_{x,n}(s_0(\tau,b,x)|\tau,b)}{dc}\bigg|\\=\;&o_p\big(n^{-1/2}\big)\end{aligned}\tag{E.49}$$

for $\bar{s}(\tau,b,x)$ a convex combination of $s_0(\tau,b,x)$ and $\hat{s}_0(\tau,b,x)$. Also note that Lemma E.1 implies

$$\begin{aligned}&\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\bigg|\frac{d\tilde{Q}^s_{x,n}(s_0(\tau,b,x)|\tau,b)}{dc}\bigg|\\&=\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\bigg|2f_{y|1,x}\big(s_0(\tau,b,x)\big)p(x)P(X_i=x)\\&\quad\times\bigg\{\frac{1}{n}\sum_{i=1}^n G_{x,i}\big(s_0(\tau,b,x)\big)+R_n(\tau,b)\bigg\}\bigg|\\&=O_p\big(n^{-1/2}\big).\end{aligned}\tag{E.50}$$

In addition, by (E.42), the mean value theorem, and $f_{y|1,x}(c)$ being uniformly bounded, we conclude that

$$\begin{aligned}&\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\bigg|\frac{d^2\tilde{Q}_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2}-2f^2_{y|1,x}\big(s_0(\tau,b,x)\big)p^2(x)P^2(X_i=1)\bigg|\\&\lesssim\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\big|f^2_{y|1,x}\big(\bar{s}(\tau,b,x)\big)-f^2_{y|1,x}\big(s_0(\tau,b,x)\big)\big|+o_p(1)\\&\lesssim\sup_{c\in\mathbf{R}}\big|f'_{y|1,x}(c)\big|\times\|\bar{s}-s_0\|_\infty+o_p(1).\end{aligned}\tag{E.51}$$

Since by assumption, $f_{y|1,x}(s_0(\tau,b,x))p(x)$ is bounded away from zero uniformly in $(\tau,b,x)\in\mathcal{S}_\varepsilon\times\mathcal{X}$, it follows from (E.51) and $\|\bar{s}-s_0\|_\infty=o_p(1)$ by Lemma E.3 that for some $\delta>0$,

$$\inf_{x\in\mathcal{X}}\inf_{(\tau,b)\in\mathcal{S}_\varepsilon}\frac{d^2\tilde{Q}_{x,n}(\bar{s}(\tau,b,x)|\tau,b)}{dc^2}>\delta\tag{E.52}$$

with probability approaching 1. Therefore, we conclude from results (E.49), (E.50), and (E.52) that we must have $\|\hat{s}_0-s_0\|_\infty=O_p(n^{-1/2})$. Hence, by (E.49) and (E.51) we conclude that

$$\sup_{x\in\mathcal{X}}\sup_{(\tau,b)\in\mathcal{S}_\varepsilon}\Big|2\big(\hat{s}_0(\tau,b,x)-s_0(\tau,b,x)\big)f^2_{y|1,x}\big(s_0(\tau,b,x)\big)p^2(x)P^2(X_i=1)$$
$$+\frac{d\tilde{Q}^s_{x,n}(s_0(\tau,b,x)|\tau,b)}{dc}\Big| \tag{E.53}$$
$$=o_p\big(n^{-1/2}\big).$$

The claim of the lemma is then established by (E.50), (E.52), and (E.53). □

LEMMA E.6. *Let Assumptions* 2.1 *and* 4.1(ii) *and* (iii) *hold, and let* W_i *be independent of* (Y_i, X_i, D_i) *with* $E[W_i]=1$, $E[W_i^2]<\infty$ *and positive a.s. Let* $\mathcal{S}_\varepsilon \equiv \{(\tau,b)\in[0,1]^2 : b\{1-p(x)\}+\varepsilon\le\tau\le p(x)+b\{1-p(x)\}-\varepsilon\ \forall x\in\mathcal{X}\}$ *for some* ε *satisfying* $0<2\varepsilon<\inf_{x\in\mathcal{X}}p(x)$ *and for some* $x_0\in\mathcal{X}$, *denote the minimizers*

$$s_0(\tau,b,x_0)=\arg\min_{c\in\mathbf{R}}Q_{x_0}(c|\tau,b).$$

If $\inf_{(\tau,b)\in\mathcal{S}_\varepsilon}f_{y|1,x}(s_0(\tau,b,x_0))p(x_0)>0$ *and* $\{Y_iD_i, X_i, D_i, W_i\}$ *is i.i.d., then the following class is Donsker*:

$$\mathcal{F}\equiv\Bigg\{f_{\tau,b}(y,x,d,w)$$
$$=\frac{w1\{y\le s_0(\tau,b,x_0),d=1,x=x_0\}+bw1\{d=0,x=x_0\}-\tau w1\{x=x_0\}}{P(X_i=x_0)p(x_0)f_{y|1,x}(s_0(\tau,b,x_0))}:$$
$$(\tau,b)\in\mathcal{S}_\varepsilon\Bigg\}.$$

PROOF. For $\delta>0$, let $\{B_j\}$ be a collection of closed balls in $\mathbf{R}^2$ with diameter δ covering $\mathcal{S}_\varepsilon$. Further notice that since $\mathcal{S}_\varepsilon\subseteq[0,1]^2$, we may select $\{B_j\}$ so its cardinality is less than $4/\delta^2$. On each B_j, define

$$\begin{aligned}
&\underline{\tau}_j=\min_{(\tau,b)\in\mathcal{S}_\varepsilon\cap B_j}\tau, && \bar{\tau}_j=\max_{(\tau,b)\in\mathcal{S}_\varepsilon\cap B_j}\tau,\\
&\underline{b}_j=\min_{(\tau,b)\in\mathcal{S}_\varepsilon\cap B_j}b, && \bar{b}_j=\max_{(\tau,b)\in\mathcal{S}_\varepsilon\cap B_j}b,\\
&\underline{s}_j=\min_{(\tau,b)\in\mathcal{S}_\varepsilon\cap B_j}s_0(\tau,b,x_0), && \bar{s}_j=\max_{(\tau,b)\in\mathcal{S}_\varepsilon\cap B_j}s_0(\tau,b,x_0),\\
&\underline{f}_j=\min_{(\tau,b)\in\mathcal{S}_\varepsilon\cap B_j}f_{y|1,x}\big(s_0(\tau,b,x_0)\big), && \bar{f}_j=\max_{(\tau,b)\in\mathcal{S}_\varepsilon\cap B_j}f_{y|1,x}\big(s_0(\tau,b,x_0)\big),
\end{aligned}\tag{E.54}$$

where we note that all minima and maxima are attained due to compactness of $\mathcal{S}_\varepsilon\cap B_j$, continuity of $s_0(\tau,b,x_0)$ by (E.5) and the implicit function theorem, and continuity of

$f_{y|1,x}(c)$ by Assumption 4.1(iii). Next, for $1 \leq j \leq \#\{B_j\}$, define the functions

$$
\begin{aligned}
l_j(y, x, d, w) \equiv & \frac{w1\{y \leq \underline{s}_j, d = 1, x = x_0\} + \underline{b}_j w1\{d = 0, x = x_0\}}{P(X_i = x_0)p(x_0)\bar{f}_j} \\
& - \frac{\bar{\tau}_j w1\{x = x_0\}}{P(X_i = x_0)p(x_0)\underline{f}_j},
\end{aligned} \tag{E.55}
$$

$$
\begin{aligned}
u_j(y, x, d, w) \equiv & \frac{w1\{y \leq \bar{s}_j, d = 1, x = x_0\} + \bar{b}_j w1\{d = 0, x = x_0\}}{P(X_i = x_0)p(x_0)\underline{f}_j} \\
& - \frac{\underline{\tau}_j w1\{x = x_0\}}{P(X_i = x_0)p(x_0)\bar{f}_j}
\end{aligned} \tag{E.56}
$$

and note that the brackets $[l_j, u_j]$ cover the class $\mathcal{F}$. Since

$$
\bar{f}_j^{-1} \leq \underline{f}_j^{-1} \leq \Big[\inf_{(\tau,b)\in\mathcal{S}_\varepsilon} f_{y|1,x}\big(s_0(\tau, b, x_0)\big)\Big]^{-1} < \infty
$$

for all j, there is a finite constant M that does not depend on j so that $M > 3E[W_i^2] \times P^{-2}(X_i = x_0)p^{-2}(x_0)\underline{f}_j^{-2}\bar{f}_j^{-2}$ uniformly in j. To bound the norm of the bracket $[l_j, u_j]$, note that for such a constant M, it follows that

$$
\begin{aligned}
& E\big[\big(u_j(Y_i, X_i, D_i, W_i) - l_j(Y_i, X_i, D_i, W_i)\big)^2\big] \\
& \quad \leq M \times (\bar{b}_j\bar{f}_j - \underline{b}_j\underline{f}_j)^2 + M \times (\bar{\tau}_j\bar{f}_j - \underline{\tau}_j\underline{f}_j)^2 \\
& \qquad + M \times E\big[\big(1\{Y_i \leq \underline{s}_j, D_i = 1, X_i = x_0\}\underline{f}_j - 1\{Y_i \leq \bar{s}_j, D_i = 1, X_i = x_0\}\bar{f}_j\big)^2\big].
\end{aligned} \tag{E.57}
$$

Next observe that by the implicit function theorem and result (E.5), we can conclude that for any $(\tau, b) \in \mathcal{S}_\varepsilon$,

$$
\begin{aligned}
\frac{ds_0(\tau, b, x_0)}{d\tau} &= \frac{1}{f_{y|1,x}(s_0(\tau, b, x_0))}, \\
\frac{ds_0(\tau, b, x_0)}{db} &= -\frac{1 - p(x_0)}{f_{y|1,x}(s_0(\tau, b, x_0))}.
\end{aligned} \tag{E.58}
$$

Since the minima and maxima in (E.54) are attained, it follows that for some (τ_1, b_1), $(\tau_2, b_2) \in B_j \cap \mathcal{S}_\varepsilon$, we have $s_0(\tau_1, b_1, x_0) = \bar{s}_j$ and $s_0(\tau_2, b_2, x_0) = \underline{s}_j$. Hence, the mean value theorem and (E.58) imply

$$
\begin{aligned}
|\bar{s}_j - \underline{s}_j| &= \left|\frac{1}{f_{y|1,x}(s_0(\tilde{\tau}, \tilde{b}, x_0))}(\tau_1 - \tau_2) + \frac{1 - p(x)}{f_{y|1,x}(s_0(\tilde{\tau}, \tilde{b}, x_0))}(b_1 - b_2)\right| \\
&\leq \frac{\sqrt{2}\delta}{\inf_{(\tau,b)\in\mathcal{S}_\varepsilon} f_{y|1,x}(s_0(\tau, b, x_0))},
\end{aligned} \tag{E.59}
$$

where $(\tilde{\tau}, \tilde{b})$ is between (τ_1, b_1) and (τ_2, b_2), and the final inequality follows by $(\tilde{\tau}, \tilde{b}) \in \mathcal{S}_\varepsilon$ by convexity of $\mathcal{S}_\varepsilon$, $(\tau_1, b_1), (\tau_2, b_2) \in B_j \cap \mathcal{S}_\varepsilon$, and B_j having diameter δ. By similar arguments and (E.59), we conclude that

$$
\begin{aligned}
|\bar{f}_j - \underline{f}_j| &\leq \sup_{c\in\mathbf{R}} \bigl|f'_{y|1,x}(c)\bigr| \times |\bar{s}_j - \underline{s}_j| \\
&\leq \sup_{c\in\mathbf{R}} \bigl|f'_{y|1,x}(c)\bigr| \times \frac{\sqrt{2}\delta}{\inf_{(\tau,b)\in\mathcal{S}_\varepsilon} f_{y|1,x}(s_0(\tau,b,x_0))}.
\end{aligned}
\tag{E.60}
$$

Since $\underline{b}_j \leq \bar{b}_j \leq 1$ due to $\bar{b}_j \in [0,1]$ and $|\bar{b}_j - \underline{b}_j| \leq \delta$ by B_j having diameter δ, we further obtain that

$$
\begin{aligned}
(\bar{b}_j\bar{f}_j - \underline{b}_j\underline{f}_j)^2 &\leq 2\bar{f}_j^2(\bar{b}_j - \underline{b}_j)^2 + 2\underline{b}_j^2(\bar{f}_j - \underline{f}_j)^2 \\
&\leq 2\sup_{c\in\mathbf{R}} f^2_{y|1,x}(c) \times \delta^2 + \frac{4\delta^2}{\inf_{(\tau,b)\in\mathcal{S}_\varepsilon} f^2_{y|1,x}(s_0(\tau,b,x_0))},
\end{aligned}
\tag{E.61}
$$

where in the final inequality, we have used result (E.60). By similar arguments, we also obtain

$$
(\bar{\tau}_j\bar{f}_j - \underline{\tau}_j\underline{f}_j)^2 \leq 2\sup_{c\in\mathbf{R}} f^2_{y|1,x}(c) \times \delta^2 + \frac{4\delta^2}{\inf_{(\tau,b)\in\mathcal{S}_\varepsilon} f^2_{y|1,x}(s_0(\tau,b,x_0))}. \tag{E.62}
$$

Also note that by direct calculation, the mean value theorem, and results (E.59) and (E.60), it follows that

$$
\begin{aligned}
&E\Bigl[\bigl(1\{Y_i \leq \underline{s}_j, D_i = 1, X_i = x_0\}\underline{f}_j - 1\{Y_i \leq \bar{s}_j, D_i = 1, X_i = x_0\}\bar{f}_j\bigr)^2\Bigr] \\
&\leq 2(\bar{f}_j - \underline{f}_j)^2 + \sup_{c\in\mathbf{R}} f^2_{y|1,x}(c) \times P(X_i = x_0)p(x_0)\bigl(F_{y|1,x}(\bar{s}_j) - F_{y|1,x}(\underline{s}_j)\bigr) \\
&\leq \frac{4\delta^2}{\inf_{(\tau,b)\in\mathcal{S}_\varepsilon} f^2_{y|1,x}(s_0(\tau,b,x_0))} + \sup_{c\in\mathbf{R}} f^3_{y|1,x}(c) \times \frac{\sqrt{2}\delta}{\inf_{(\tau,b)\in\mathcal{S}_\varepsilon} f_{y|1,x}(s_0(\tau,b,x_0))}.
\end{aligned}
\tag{E.63}
$$

Thus, from (E.57), (E.61), and (E.62), it follows that for $\delta < 1$ and some constant K not depending on j,

$$
E\Bigl[\bigl(u_j(Y_i,X_i,D_i,W_i) - l_j(Y_i,X_i,D_i,W_i)\bigr)^2\Bigr] \leq K\delta. \tag{E.64}
$$

Since $\#\{B_j\} \leq 4/\delta^2$, we can, therefore, conclude that $N_{[]}(\delta,\mathcal{F},\|\cdot\|_{L^2}) \leq 4K^2/\delta^2$ and, hence, by Theorem 2.5.6 in van der Vaart and Wellner (1996), it follows that the class $\mathcal{F}$ is Donsker. □

Lemma E.7. *Let Assumptions* 2.1 *and* 4.1(ii) *and* (iii) *hold, and let W_i be independent of* (Y_i, X_i, D_i) *with* $E[W_i] = 1$, $E[W_i^2] < \infty$, *positive a.s.*, $\mathcal{S}_\zeta \equiv \{(\tau,b) \in [0,1]^2 : b\{1 - p(x)\} +$

$\zeta \le \tau \le p(x) + b\{1 - p(x)\} - \zeta \ \forall x \in \mathcal{X}\}$, *and*

$$\tilde{p}(x) \equiv \frac{\sum_{i=1}^{n} W_i 1\{D_i = 1, X_i = x\}}{\sum_{i=1}^{n} W_i 1\{X_i = x\}}, \qquad p(x) \equiv P(D_i = 1 | X_i = x),$$

$$s_0(\tau, b, x) = \arg\min_{c \in \mathbf{R}} Q_x(c|\tau, b).$$

If $\inf_{(\tau,b,x)\in\mathcal{S}_\zeta\times\mathcal{X}} f_{y|1,x}(s_0(\tau,b,x))p(x) > 0$ *and* $\{Y_i D_i, X_i, D_i, W_i\}$ *is an i.i.d. sample, then for* $a \in \{-1, 1\}$,

$$\begin{aligned} & s_0\big(\tau, \tau + ak\tilde{p}(x), x\big) - s_0\big(\tau, \tau + akp(x), x\big) \\ &= -\frac{(1-p(x))ka}{f_{y|1,x}(s_0(\tau, \tau + akp(x), x))P(X = x)} \times \frac{1}{n}\sum_{i=1}^{n} R(X_i, W_i, x) + o_p\big(n^{-1/2}\big), \end{aligned} \tag{E.65}$$

where $R(W_i, X_i, x) = p(x)\{P(X = x) - W_i 1\{X_i = x\}\} + W_i 1\{D_i = 1, X_i = x\} - P(D = 1, X = x)$ *and* (E.65) *holds uniformly in* $(\mathcal{B}_\zeta \times \mathcal{X})$. *Moreover, the right hand side of* (E.65) *is Donsker.*

Proof. First observe that $(\tau, k) \in \mathcal{B}_\zeta$ implies $(\tau, \tau + akp(x)) \in \mathcal{S}_\zeta$ for all $x \in \mathcal{X}$ and that with probability tending to 1, $(\tau, \tau + ak\tilde{p}(x)) \in \mathcal{S}_\zeta$ for all $(\tau, k) \in \mathcal{B}_\zeta$. In addition, also note that

$$\tilde{p}(x) - p(x) = \frac{1}{nP(X = x)}\sum_{i=1}^{n} R(X_i, W_i, x) + o_p\big(n^{-1/2}\big) \tag{E.66}$$

by an application of the Delta method and $\inf_{x\in\mathcal{X}} P(X = x) > 0$ due to X having finite support. Moreover, by the mean value theorem and (E.58), we obtain for some $\bar{b}(\tau, k)$ between $\tau + ak\tilde{p}(x)$ and $\tau + akp(x)$ that

$$\begin{aligned} & s_0\big(\tau, \tau + ak\tilde{p}(x), x\big) - s_0\big(\tau, \tau + akp(x), x\big) \\ &= -\frac{(1-p(x))ka}{f_{y|1,x}(s_0(\tau, \bar{b}(\tau,k), x))}\big(\tilde{p}(x) - p(x)\big) \\ &= -\frac{(1-p(x))ka}{f_{y|1,x}(s_0(\tau, \tau + akp(x), x))}\big(\tilde{p}(x) - p(x)\big) + o_p\big(n^{-1/2}\big), \end{aligned} \tag{E.67}$$

where the second equality follows from $(\tau, \bar{b}(\tau, k)) \in \mathcal{S}_\zeta$ for all (τ, k) with probability approaching 1 by convexity of $\mathcal{S}_\zeta$, $\inf_{(\tau,b,x)\in\mathcal{S}_\zeta\times\mathcal{X}} f_{y|1,x}(s_0(\tau,b,x))p(x) > 0$, and $\sup_{(\tau,k)\in\mathcal{B}_\zeta} |ak(\tilde{p}(x) - p(x))| = o_p(1)$ uniformly in $\mathcal{X}$. The first claim of the lemma then follows by combining (E.66) and (E.67).

Finally, observe that the right hand side of (E.65) is trivially Donsker since $R(X_i, W_i, x)$ does not depend on (k, τ) and the function $(1 - p(x))ka/(f_{y|1,x}(s_0(\tau, \tau + akp(x),$

$x))P(X=x))$ is uniformly continuous on $(\tau,k)\in\mathcal{B}_\zeta$ due to $\inf_{(\tau,b,x)\in\mathcal{S}_\zeta\times\mathcal{X}} f_{y|1,x}(s_0(\tau,b,x))p(x)>0$. □

Proof of Theorem 4.1. Throughout the proof, we exploit Lemmas E.5 and E.6 applied with $W_i=1$ with probability 1, so that $\tilde{Q}_{x,n}(c|\tau,b)=Q_{x,n}(c|\tau,b)$ for all (τ,b) in $\mathcal{S}_\zeta$, where

$$\begin{aligned}\mathcal{S}_\zeta\equiv\Big\{&(\tau,b)\in[0,1]^2:\\ &b\big\{1-p(x)\big\}+\zeta\le\tau\le p(x)+b\big\{1-p(x)\big\}-\zeta\ \forall x\in\mathcal{X}\Big\}.\end{aligned}\tag{E.68}$$

Also notice that for every $(\tau,k)\in\mathcal{B}_\zeta$ and all $x\in\mathcal{X}$, the points $(\tau,\tau+kp(x)),(\tau,\tau-kp(x))\in\mathcal{S}_\zeta$, while with probability approaching 1, $(\tau,\tau+k\hat{p}(x))$ and $(\tau,\tau-k\hat{p}(x))$ also belong to $\mathcal{S}_\zeta$. Therefore, for $s_0(\tau,b,x)$ and $\hat{s}_0(\tau,b,x)$ as defined in (E.47), we obtain from Lemmas E.5 and E.6, applied with $W_i=1$ a.s., that

$$\begin{aligned}&\Big|\Big(\hat{s}_0\big(\tau,\tau+akp(x),x\big)-s_0\big(\tau,\tau+akp(x),x\big)\Big)\\ &\qquad-\Big(\hat{s}_0\big(\tau,\tau+ak\hat{p}(x),x\big)-s_0\big(\tau,\tau+ak\hat{p}(x),x\big)\Big)\Big|\\ &\quad=o_p\big(n^{-1/2}\big)\end{aligned}\tag{E.69}$$

uniformly in $(\tau,k,x)\in\mathcal{B}_\zeta\times\mathcal{X}$ and $a\in\{-1,1\}$. Moreover, by Lemma E.7 applied with $W_i=1$ a.s.,

$$\begin{aligned}&s_0\big(\tau,\tau+ak\hat{p}(x),x\big)-s_0\big(\tau,\tau+akp(x),x\big)\\ &\quad=-\frac{(1-p(x))ka}{f_{y|1,x}(s_0(\tau,\tau+akp(x),x))P(X=x)}\times\frac{1}{n}\sum_{i=1}^n R(X_i,x)+o_p\big(n^{-1/2}\big),\end{aligned}\tag{E.70}$$

where $R(X_i,x)=p(x)\{P(X=x)-1\{X_i=x\}\}+1\{D_i=1,X_i=x\}-P(D=1,X=x)$ again uniformly in $(\tau,k,x)\in\mathcal{B}_\zeta\times\mathcal{X}$. Also observe that since $(\tau,\tau+k\hat{p}(x))$ and $(\tau,\tau-k\hat{p}(x))$ belong to $\mathcal{S}_\zeta$ with probability approaching 1, we obtain uniformly in $(\tau,k,x)\in\mathcal{B}_\zeta\times\mathcal{X}$ that

$$\begin{aligned}q_L(\tau,k|x)&=s_0\big(\tau,\tau+kp(x),x\big),\\ q_U(\tau,k|x)&=s_0\big(\tau,\tau-kp(x),x\big),\\ \hat{q}_L(\tau,k|x)&=\hat{s}_0\big(\tau,\tau+k\hat{p}(x),x\big)+o_p\big(n^{-1/2}\big),\\ \hat{q}_U(\tau,k|x)&=\hat{s}_0\big(\tau,\tau-k\hat{p}(x),x\big)+o_p\big(n^{-1/2}\big).\end{aligned}\tag{E.71}$$

Therefore, combining results (E.69)–(E.71), and exploiting Lemmas E.5, E.6, and E.7 and the sum of Donsker classes being Donsker, we conclude that for J a Gaussian process on $L^\infty(\mathcal{B}_\zeta\times\mathcal{X})\times L^\infty(\mathcal{B}_\zeta\times\mathcal{X})$,

$$\sqrt{n}C_n\xrightarrow{L}J,\quad C_n(\tau,k,x)\equiv\begin{pmatrix}\hat{q}_L(\tau,k|x)-q_L(\tau,k|x)\\ \hat{q}_U(\tau,k|x)-q_U(\tau,k|x)\end{pmatrix}.\tag{E.72}$$

To establish the second claim of the theorem, observe that since X has finite support, we may denote $\mathcal{X} = \{x_1, \dots, x_{|\mathcal{X}|}\}$ and define the matrix $B = (P(X_i = x_1)x_1, \dots, P(X_i = x_{|\mathcal{X}|})x_{|\mathcal{X}|})$ as well as the vector

$$w \equiv \lambda'\big(E_S[X_i X_i']\big)^{-1} B. \tag{E.73}$$

Since w is also a function on $\mathcal{X}$, we refer to its coordinates by $w(x)$. Solving the linear programming problems defined in equations (24) and (25), we obtain the closed form solution

$$\begin{aligned}\pi_L(\tau, k) &= \sum_{x\in\mathcal{X}} \big\{1\{w(x) \geq 0\}w(x)q_L(\tau, k|x) + 1\{w(x) \leq 0\}w(x)q_U(\tau, k|x)\big\},\\ \pi_U(\tau, k) &= \sum_{x\in\mathcal{X}} \big\{1\{w(x) \geq 0\}w(x)q_U(\tau, k|x) + 1\{w(x) \leq 0\}w(x)q_L(\tau, k|x)\big\},\end{aligned} \tag{E.74}$$

with a similar representation holding for $(\hat{\pi}_L(\tau, k), \hat{\pi}_U(\tau, k))$, but with $(\hat{q}_L(\tau, k|x), \hat{q}_U(\tau, k|x))$ in place of $(q_L(\tau, k|x), q_U(\tau, k|x))$. We hence define the linear map $K: L^\infty(\mathcal{B}_\zeta \times \mathcal{X}) \times L^\infty(\mathcal{B}_\zeta \times \mathcal{X}) \to L^\infty(\mathcal{B}_\zeta) \times L^\infty(\mathcal{B}_\zeta)$, to be given by

$$K(\theta)(\tau, k) \equiv \begin{pmatrix} \displaystyle\sum_{x\in\mathcal{X}} \big\{1\{w(x) \geq 0\}w(x)\theta^{(1)}(\tau, k, x) \\ \qquad + 1\{w(x) \leq 0\}w(x)\theta^{(2)}(\tau, k, x)\big\} \\ \displaystyle\sum_{x\in\mathcal{X}} \big\{1\{w(x) \geq 0\}w(x)\theta^{(2)}(\tau, k, x) \\ \qquad + 1\{w(x) \leq 0\}w(x)\theta^{(1)}(\tau, k, x)\big\}\end{pmatrix}, \tag{E.75}$$

where for any $\theta \in L^\infty(\mathcal{X} \times \mathcal{B}_\zeta) \times L^\infty(\mathcal{B}_\zeta \times \mathcal{X})$, $\theta^{(i)}(\tau, k, x)$ denotes the ith coordinate of the two dimensional vector $\theta(\tau, k, x)$. It then follows from (E.72), (E.74), and (E.75) that

$$\sqrt{n}\begin{pmatrix}\hat{\pi}_L - \pi_L \\ \hat{\pi}_U - \pi_U\end{pmatrix} = \sqrt{n}K(C_n). \tag{E.76}$$

Moreover, employing the norm $\|\cdot\|_\infty + \|\cdot\|_\infty$ on the product spaces $L^\infty(\mathcal{B}_\zeta \times \mathcal{X}) \times L^\infty(\mathcal{B}_\zeta \times \mathcal{X})$ and $L^\infty(\mathcal{B}_\zeta) \times L^\infty(\mathcal{B}_\zeta)$, we can then obtain by direct calculation that for any $\theta \in L^\infty(\mathcal{B}_\zeta \times \mathcal{X}) \times L^\infty(\mathcal{B}_\zeta \times \mathcal{X})$,

$$\big\|K(\theta)\big\|_\infty \leq 2\sum_{x\in\mathcal{X}} \big|w(x)\big| \times \sup_{x\in\mathcal{X}} \sup_{(\tau,b)\in\mathcal{S}_\zeta} \big|\theta(\tau, b, x)\big| = 2\sum_{x\in\mathcal{X}}\big|w(x)\big| \times \|\theta\|_\infty, \tag{E.77}$$

which implies the linear map K is continuous. Therefore, the theorem is established by (E.72), (E.76), the linearity of K, and the continuous mapping theorem. □

Proof of Theorem 4.2. For a metric space $\mathbb{D}$, let $\mathrm{BL}_c(\mathbb{D})$ denote the set of real valued bounded Lipschitz functions with supremum norm and Lipschitz constant less than or equal to c. We first aim to show that

$$\sup_{h\in\mathrm{BL}_1(\mathbf{R})} \big|E\big[h\big(L(\tilde{G}_\omega)\big)|\mathcal{Z}_n\big] - E\big[h\big(L(G_\omega)\big)\big]\big| = o_p(1), \tag{E.78}$$

where $\mathcal{Z}_n = \{Y_i D_i, X_i, D_i\}_{i=1}^n$ and $E[h(\tilde{Z})|\mathcal{Z}_n]$ denotes outer expectation over $\{W_i\}_{i=1}^n$ with $\mathcal{Z}_n$ fixed. Let

$$\hat{s}_0(\tau, b, x) \in \arg\min_{c\in\mathbf{R}} Q_{x,n}(c|\tau, b), \qquad \tilde{s}_0(\tau, b, x) \in \arg\min_{c\in\mathbf{R}} \tilde{Q}_{x,n}(c|\tau, b),$$
$$s_0(\tau, b, x) \in \arg\min_{c\in\mathbf{R}} Q_x(c|\tau, b). \tag{E.79}$$

Also note that with probability approaching 1 the points $(\tau, \tau + ak\tilde{p}(x)) \in \mathcal{S}_\zeta$ for all $(\tau, k, x) \in \mathcal{B}_\zeta \times \mathcal{X}$ and $a \in \{-1, 1\}$ for $\mathcal{S}_\zeta$ as in (E.68). Hence, arguing as in (E.69) and (E.70), we obtain

$$\begin{aligned}&\tilde{q}_L(\tau, k|x) - \hat{q}_L(\tau, k|x)\\ &\quad = \tilde{s}_0\big(\tau, \tau + kp(x), x\big) - \hat{s}_0\big(\tau, \tau + kp(x), x\big)\\ &\qquad - \frac{(1-p(x))k}{f_{y|1,x}(s_0(\tau, \tau + kp(x), x))P(X = x)} \times \frac{1}{n}\sum_{i=1}^n \Delta R(X_i, W_i, x) + o_p\big(n^{-1/2}\big),\end{aligned} \tag{E.80}$$

$$\begin{aligned}&\tilde{q}_U(\tau, k|x) - \hat{q}_U(\tau, k|x)\\ &\quad = \tilde{s}_0\big(\tau, \tau - kp(x), x\big) - \hat{s}_0\big(\tau, \tau - kp(x), x\big)\\ &\qquad + \frac{(1-p(x))k}{f_{y|1,x}(s_0(\tau, \tau - kp(x), x))P(X = x)} \times \frac{1}{n}\sum_{i=1}^n \Delta R(X_i, W_i, x) + o_p\big(n^{-1/2}\big),\end{aligned} \tag{E.81}$$

where $\Delta R(X_i, W_i, x) = (1 - W_i)(1\{X_i = x\}p(x) - 1\{D_i = 1, X_i = x\})$ and both statements hold uniformly in $(\tau, k, x) \in \mathcal{B}_\zeta \times \mathcal{X}$. Also note that for the operator K as defined in (E.75), we have

$$\sqrt{n}\begin{pmatrix}\tilde{\pi}_L - \hat{\pi}_L\\ \tilde{\pi}_U - \hat{\pi}_U\end{pmatrix} = \sqrt{n}K(\tilde{C}_n), \quad \tilde{C}_n(\tau, k, x) \equiv \begin{pmatrix}\tilde{q}_L(\tau, k|x) - \hat{q}_L(\tau, k|x)\\ \tilde{q}_U(\tau, k|x) - \hat{q}_U(\tau, k|x)\end{pmatrix}. \tag{E.82}$$

By Lemmas E.5, E.6, and E.7, results (E.80) and (E.81), and Theorem 2.9.2 in van der Vaart and Wellner (1996), the process $\sqrt{n}\tilde{C}_n$ converges unconditionally to a tight Gaussian process on $L^\infty(\mathcal{B}_\zeta \times \mathcal{X})$. Hence, by the continuous mapping theorem, $\sqrt{n}K(\tilde{C}_n)$ is asymptotically tight. Define

$$\bar{G}_\omega \equiv \sqrt{n}\begin{pmatrix}(\tilde{\pi}_L - \hat{\pi}_L)/\omega_L\\ (\tilde{\pi}_U - \hat{\pi}_U)/\omega_U\end{pmatrix}, \tag{E.83}$$

and notice that $\omega_L(\tau, k)$ and $\omega_U(\tau, k)$ being bounded away from zero, $\hat{\omega}_L(\tau, k)$ and $\hat{\omega}_U(\tau, k)$ being uniformly consistent by Assumption 4.2(ii), and $\sqrt{n}K(\tilde{C}_n)$ being asymptotically tight imply that

$$\begin{aligned}&\big|L(\tilde{G}_\omega) - L(\bar{G}_\omega)\big|\\ &\quad \le \sup_{(\tau,k)\in\mathcal{B}_\zeta} M_0\left|\frac{\sqrt{n}(\hat{\pi}_L(\tau, k) - \tilde{\pi}_L(\tau, k))}{\hat{\omega}_L(\tau, k)} - \frac{\sqrt{n}(\hat{\pi}_L(\tau, k) - \tilde{\pi}_L(\tau, k))}{\omega_L(\tau, k)}\right|\end{aligned} \tag{E.84}$$

$$+\sup_{(\tau,k)\in\mathcal{B}_\zeta} M_0\left|\frac{\sqrt{n}(\hat{\pi}_U(\tau,k)-\tilde{\pi}_U(\tau,k))}{\hat{\omega}_U(\tau,k)}-\frac{\sqrt{n}(\hat{\pi}_U(\tau,k)-\tilde{\pi}_U(\tau,k))}{\omega_U(\tau,k)}\right|$$
$$=o_p(1)$$

for some constant M_0 due to L being Lipschitz. By definition of BL_1, all $h\in\mathrm{BL}_1$ have Lipschitz constant less than or equal to 1 and are also bounded by 1. Hence, for any $\eta>0$, Markov's inequality implies

$$\begin{aligned}
&P\Big(\sup_{h\in\mathrm{BL}_1(\mathbf{R})}\big|E\big[h\big(L(\tilde{G}_\omega)\big)|\mathcal{Z}_n\big]-E\big[h\big(L(\bar{G}_\omega)\big)|\mathcal{Z}_n\big]\big|>\eta\Big)\\
&\le P\bigg(2P\Big(\big|L(\tilde{G}_\omega)-L(\bar{G}_\omega)\big|>\frac{\eta}{2}\Big|\mathcal{Z}_n\Big)\\
&\qquad+\frac{\eta}{2}P\Big(\big|L(\tilde{G}_\omega)-L(\bar{G}_\omega)\big|\le\frac{\eta}{2}\Big|\mathcal{Z}_n\Big)>\eta\bigg)\\
&\le\frac{4}{\eta}E\Big[E\Big[1\Big\{\big|L(\tilde{G}_\omega)-L(\bar{G}_\omega)\big|>\frac{\eta}{2}\Big\}\Big|\mathcal{Z}_n\Big]\Big].
\end{aligned}\tag{E.85}$$

Therefore, by (E.84), (E.85), and Lemma 1.2.6 in van der Vaart and Wellner (1996), we obtain

$$\begin{aligned}
&P\Big(\sup_{h\in\mathrm{BL}_1(\mathbf{R})}\big|E\big[h\big(L(\tilde{G}_\omega)\big)|\mathcal{Z}_n\big]-E\big[h\big(L(\bar{G}_\omega)\big)|\mathcal{Z}_n\big]\big|>\eta\Big)\\
&\le\frac{4}{\eta}P\Big(\big|L(\tilde{G}_\omega)-L(\bar{G}_\omega)\big|>\frac{\eta}{2}\Big)=o(1).
\end{aligned}\tag{E.86}$$

Next, let $\overset{L}{=}$ stands for "equal in law" and notice that for J the Gaussian process in (E.72),

$$L(G_\omega)\overset{L}{=}T\circ K(J),\qquad L(\bar{G}_\omega)=\sqrt{n}L\circ K(\tilde{C}_n)\tag{E.87}$$

due to the continuous mapping theorem and (E.82). For $w(x)$ as defined in (E.70) and $C_0\equiv 2\sum_{x\in\mathcal{X}}|w(x)|$, it follows from linearity of K and (E.75), that K is Lipschitz with Lipschitz constant C_0. Therefore, for any $h\in\mathrm{BL}_1(\mathbf{R})$, result (E.87) implies that $h\circ L\circ K\in\mathrm{BL}_{C_0M_0}(L^\infty(\mathcal{B}_\zeta\times\mathcal{X}))$ for some $M_0>0$ and, hence,

$$\begin{aligned}
&\sup_{h\in\mathrm{BL}_1(\mathbf{R})}\big|E\big[h\big(L(\bar{G}_\omega)\big)|\mathcal{Z}_n\big]-E\big[h\big(L(G_\omega)\big)\big]\big|\\
&\le\sup_{h\in\mathrm{BL}_{C_0M_0}(L^\infty(\mathcal{B}_\zeta\times\mathcal{X}))}\big|E\big[h(\bar{G}_\omega)|\mathcal{Z}_n\big]-E\big[h(J)\big]\big|=o_p(1),
\end{aligned}\tag{E.88}$$

where the final equality follows from (E.80), (E.81), (E.87), arguing as in (E.85) and (E.86), Lemmas E.6 and E.7, and Theorem 2.9.6 in van der Vaart and Wellner (1996). Hence, (E.86) and (E.88) establish (E.78).

Next, we aim to show that for all $t\in\mathbf{R}$ at which the CDF of $L(G_\omega)$ is continuous and for any $\eta>0$,

$$P\big(\big|P\big(L(\tilde{G}_\omega)\le t|\mathcal{Z}_n\big)-P\big(L(G_\omega)\le t\big)\big|>\eta\big)=o(1).\tag{E.89}$$

Toward this end, for every $\lambda > 0$ and t at which the CDF of $L(G_\omega)$ is continuous, define the functions

$$\begin{aligned} h^U_{\lambda,t}(u) &= 1 - 1\{u > t\}\min\{\lambda(u-t), 1\}, \\ h^L_{\lambda,t}(u) &= 1\{u < t\}\min\{\lambda(t-u), 1\}. \end{aligned} \tag{E.90}$$

Notice that by construction, $h^L_{\lambda,t}(u) \le 1\{u \le t\} \le h^U_{\lambda,t}(u)$ for all $u \in \mathbf{R}$, the functions $h^L_{\lambda,t}$ and $h^U_{\lambda,t}$ are both bounded by 1, and they are both Lipschitz with Lipschitz constant λ. Also by direct calculation,

$$0 \le E\big[h^U_{\lambda,t}\big(L(G_\omega)\big) - h^L_{\lambda,t}\big(L(G_\omega)\big)\big] \le P\big(t - \lambda^{-1} \le L(G_\omega) \le t + \lambda^{-1}\big). \tag{E.91}$$

Therefore, exploiting that $h^L_{\lambda,t}, h^U_{\lambda,t} \in \mathrm{BL}_\lambda(\mathbf{R})$ and that $h \in \mathrm{BL}_\lambda(\mathbf{R})$ implies $\lambda^{-1}h \in \mathrm{BL}_1(\mathbf{R})$, we obtain

$$\begin{aligned} &\big|P\big(L(\tilde{G}_\omega) \le t|\mathcal{Z}_n\big) - P\big(L(G_\omega) \le t\big)\big| \\ &\quad \le \big|E\big[h^L_{\lambda,t}\big(L(\tilde{G}_\omega)\big)|\mathcal{Z}_n\big] - E\big[h^U_{\lambda,t}\big(L(G_\omega)\big)\big]\big| \\ &\qquad + \big|E\big[h^U_{\lambda,t}\big(L(\tilde{G}_\omega)\big)|\mathcal{Z}_n\big] - E\big[h^L_{\lambda,t}\big(L(G_\omega)\big)\big]\big| \\ &\quad \le 2 \sup_{h \in \mathrm{BL}_\lambda(\mathbf{R})} \big|E\big[h\big(L(\tilde{G}_\omega)\big)|\mathcal{Z}_n\big] - E\big[h\big(L(G_\omega)\big)\big]\big| \\ &\qquad + 2P\big(t - \lambda^{-1} \le L(G_\omega) \le t + \lambda^{-1}\big) \\ &\quad = 2\lambda \sup_{h \in \mathrm{BL}_1(\mathbf{R})} \big|E\big[h\big(L(\tilde{G}_\omega)\big)|\mathcal{Z}_n\big] - E\big[h\big(L(G_\omega)\big)\big]\big| \\ &\qquad + 2P\big(t - \lambda^{-1} \le L(G_\omega) \le t + \lambda^{-1}\big) \end{aligned} \tag{E.92}$$

for any $\lambda > 0$. Moreover, we may select a λ_η sufficiently large so that $2P(t - \lambda_\eta^{-1} \le L(G_\omega) \le t + \lambda_\eta^{-1}) < \eta/2$ due to t being a continuity point of the CDF of $L(G_\omega)$. Therefore, from (E.92), we obtain

$$\begin{aligned} &P\big(\big|P\big(L(\tilde{G}_\omega) \le t|\mathcal{Z}_n\big) - P\big(L(G_\omega) \le t\big)\big| > \eta\big) \\ &\quad \le P\bigg(2\lambda_\eta \sup_{h \in \mathrm{BL}_1(\mathbf{R})} \big|E\big[h\big(L(\tilde{G}_\omega)\big)|\mathcal{Z}_n\big] - E\big[h\big(L(G_\omega)\big)\big]\big| > \frac{\eta}{2}\bigg) = o(1), \end{aligned} \tag{E.93}$$

where the final equality follows from (E.78).

Finally, note that since the CDF of $L(G_\omega)$ is strictly increasing and continuous at $r_{1-\alpha}$, we obtain that

$$P\big(L(G_\omega) \le r_{1-\alpha} - \varepsilon\big) < 1 - \alpha < P\big(L(G_\omega) \le r_{1-\alpha} + \varepsilon\big) \tag{E.94}$$

$\forall \varepsilon > 0$. Define the event $A_n \equiv \{P(L(\tilde{G}_\omega) \le r_{1-\alpha} - \varepsilon|\mathcal{Z}_n) < 1 - \alpha < P(L(\tilde{G}_\omega) \le r_{1-\alpha} + \varepsilon|\mathcal{Z}_n)\}$ and notice that

$$P\big(|\tilde{r}_{1-\alpha} - r_{1-\alpha}| \le \varepsilon\big) \ge P(A_n) \to 1, \tag{E.95}$$

where the inequality follows by definition of $\tilde{r}_{1-\alpha}$, and the second result is implied by (E.89) and (E.94). □

References

Angrist, J., V. Chernozhukov, and I. Fernádez-Val (2006), "Quantile regression under misspecification, with an application to the U.S. wage structure." *Econometrica*, 74, 539–563. [3]

Bound, J. and A. Krueger (1991), "The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?" *Journal of Labor Economics*, 9, 1–24. [3]

Imbens, G. W. and C. F. Manski (2004), "Confidence intervals for partially identified parameters." *Econometrica*, 72, 1845–1857. [5]

van der Vaart, A. W. and J. A. Wellner (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. [9, 13, 15, 21, 25, 26]

Submitted July, 2011. Final version accepted April, 2012.