Econometrica Supplementary Material

SUPPLEMENT TO "EFFICIENT SEMIPARAMETRIC ESTIMATION OF QUANTILE TREATMENT EFFECTS"

(*Econometrica*, Vol. 75, No. 1, January, 2007, 259–276)

BY SERGIO FIRPO[1]

This supplement presents (i) detailed results from an empirical example, (ii) results of a Monte Carlo exercise, (iii) details of the variance estimation, and (iv) proofs of the theoretical results.

KEYWORDS: Quantile treatment effects, propensity score, semiparametric efficiency bounds, efficient estimation, semiparametric estimation.

1. APPLICATION TO A JOB TRAINING PROGRAM

IN THIS SECTION we present the details of an empirical application for the quantile treatment effect (QTE) estimators proposed in the main paper. This application uses a job training program data set that was first analyzed by LaLonde (1986) and later by many others, including Heckman and Hotz (1989), Dehejia and Wahba (1999), Smith and Todd (2001, 2005), and Imbens (2003).

1.1. *Data Set*

The original data set from the National Supported Work Program (NSW) is well described in LaLonde (1986). The program was designed as an experiment where applicants were randomly assigned into treatment. The treatment was work experience in a wide range of possible activities, like learning to operating a restaurant or a child care center, for a period not exceeding 12 months. Eligible participants were targeted from recipients of Aid to Families With Dependent Children, former addicts, former offenders, and young school dropouts. The NSW data set consists of information on earnings and employment (outcome variables), whether treated or not, and background characteristics, such as education, ethnicity, age, and employment variables before treatment. LaLonde used this experimental data set as a benchmark for comparisons with the case in which comparison samples come from non-experimental data sets, as for example, comparison groups based on Panel Study of Income Dynamics (PSID) and on Westat's Matched Current Population Survey–Social Security Administration file. We use only a subsample from the PSID, which corresponds to the subsample termed PSID-1 by Dehejia and Wahba (1999). Summary statistics for the two data sets are presented in Table I.

[1]I am indebted to Guido Imbens and Jim Powell for their advice, support, and many suggestions. Thanks go to Rafael Cayres, who provided excellent research assistance with Matlab programming of Monte Carlo routines. Financial support from CAPES—Brazil is acknowledged. All errors are mine.

TABLE I

SUMMARY STATISTICS

	Treatment Group Sample Size 185			Experimental Control Group Sample Size 260			Nonexperimental Comparison Group Sample Size 2,490		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Earnings (1978)	6,349.1	0.0	6,0307.9	4,554.8	0.0	39,483.5	21,553.9	0.0	121,174.0
	(7,867.4)			(5,483.8)			(15,555.3)		
Age	25.8	17.0	48.0	25.1	17.0	55.0	34.9	18.0	55.0
	(7.2)			(7.1)			(10.4)		
Education	10.3	4.0	16.0	10.1	3.0	14.0	12.1	0.0	17.0
	(2.0)			(1.6)			(3.1)		
Dropout	0.7	—	—	0.8	—	—	0.3	—	—
Black	0.8	—	—	0.8	—	—	0.3	—	—
Hispanic	0.1	—	—	0.1	—	—	0.0	—	—
Married	0.2	—	—	0.2	—	—	0.9	—	—
Earnings (1974)	2,095.6	0.0	35,040.1	2,107.0	0.0	39,570.7	19,428.8	0.0	137,149.0
	(4,886.6)			(5,687.9)			(13,406.9)		
Earnings (1975)	1,532.1	0.0	25,142.2	1,266.9	0.0	23,032.0	19,063.3	0.0	156,653.0
	(3,219.3)			(3,103.0)			(13,596.9)		
Unemployed (1974)	0.7	—	—	0.8	—	—	0.1	—	—
Unemployed (1975)	0.6	—	—	0.7	—	—	0.1	—	—

As this table reveals, the nonexperimental comparison group is essentially different from the treated group, which leads us to turn attention to the parameters of the treatment effect on the treated. In what follows, the outcome variable is earnings in 1978.[2] As in Dehejia and Wahba (1999), we consider male workers only.

LaLonde found that nonexperimental comparison samples are poor substitutes for experimental data. Some reasons for this finding are described in the survey paper by Heckman, LaLonde, and Smith (1999) and are explored subsequently by Smith and Todd (2001, 2005). The key aspects are related to the comparability between the information coming from the surveys. First, the surveys have different questions and questionnaires; second, they cover different labor markets; and third, they do not bring many compelling controlling variables. A fourth reason why nonexperimental and experimental data sets provide different answers was offered by Dehejia and Wahba (1999), who argued that some important preprogram variables should be included as control variables and that a more flexible methodology that did not rely on parametric wage regressions should be employed.[3]

1.2. *Using the QTE Estimation Method*

With the data set just described, we used treatment and observational comparison groups to provide estimates of the quantile treatment effect on the treated (QTT) and to compare them with experimental QTT estimates, which are just differences between quantiles of the treated and the experimental controls, without any reweighting. Our results are presented in Table II and Figures 1–5.

We estimated the propensity score by approximating its log-odds ratio by a polynomial. The order of such a polynomial was guided by a leave-one-out cross-validation method based on Hall (1987), in which the optimal number of terms minimizes a Kullback–Leibler distance. The order in which the polynomial terms were added was given, however, by the following method. Starting with a full linear model, we included terms according to their degree: those with lower degree entered first. Then, within the same degree, we sequentially added terms with fewer variables; thus interactions entered only after we considered all pure terms (e.g., x_1^2 and x_2^2 are added before $x_1 \cdot x_2$). This is a nonparametric extension to the propensity-score model selection discussed extensively in Dehejia and Wahba (1999) and in Rosenbaum and Rubin (1984),

[2]Earnings are measured in 1982 U.S. dollars.

[3]Dehejia and Wahba included information on the previous two years earnings, which reduced the treated sample by about 40%. See Dehejia and Wahba (1999) and Smith and Todd (2001). A possible weakness of Dehejia and Wahba's method is explored in Smith and Todd (2005), who argue that such results may not be robust to other ways of constructing the observational comparison samples.

TABLE II

QUANTILES AND QUANTILE TREATMENT EFFECTS

				Quantiles		
			Mean	0.25	0.50	0.75
Observational data	(a)	$Y(1)\|T=1$	6,349.1	485.2	4,232.3	9,643.0
			(576.9)	(329.7)	(668.3)	(951.6)
	(b)	$Y(0)\|T=1$	5,185.8	0.0	2,305.2	3,694.3
			(1,662.7)	(3.9)	(633.9)	(940.0)
	(c)	$Y(0)\|T=0$	21,554.0	11,526.0	20,688.0	29,554.0
			(311.67)	(397.2)	(344.2)	(380.9)
Experimental data	(d)	$Y(1)\|T=1$	6,349.1	485.2	4,232.3	9,643.0
			(576.9)	(329.7)	(668.3)	(951.6)
	(e)	$Y(0)\|T=0$	4,554.8	0.0	3,116.5	7,293.9
			(339.4)	(0.1)	(539.7)	(633.1)
Observational treatment effects	(a), (b)		1,163.4	485.2	1,927.1	5,948.7
			(1,735.9)	(330.5)	(1,018.3)	(1,387.3)
Experimental treatment effects	(d), (e)		1,794.3	485.2	1,115.8	2,349.1
			(669.3)	(332.1)	(859.0)	(1,142.9)
Difference in treatment effects	(b)–(e)		630.9	0.0	−811.3	−3,599.6
			(1,860.5)	(468.5)	(1,332.2)	(1,797.4)

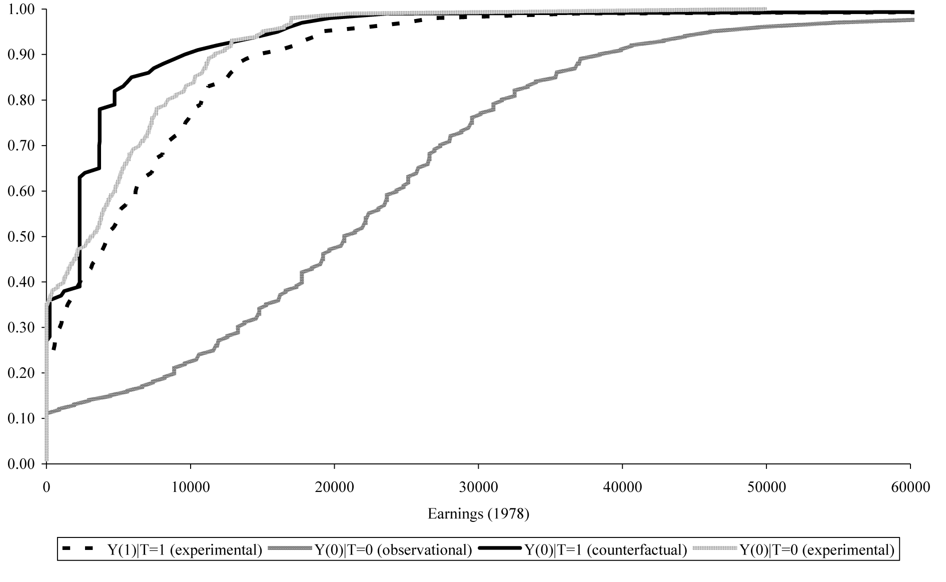

FIGURE 1.—Experimental and nonexperimental c.d.f.'s of potential outcomes.

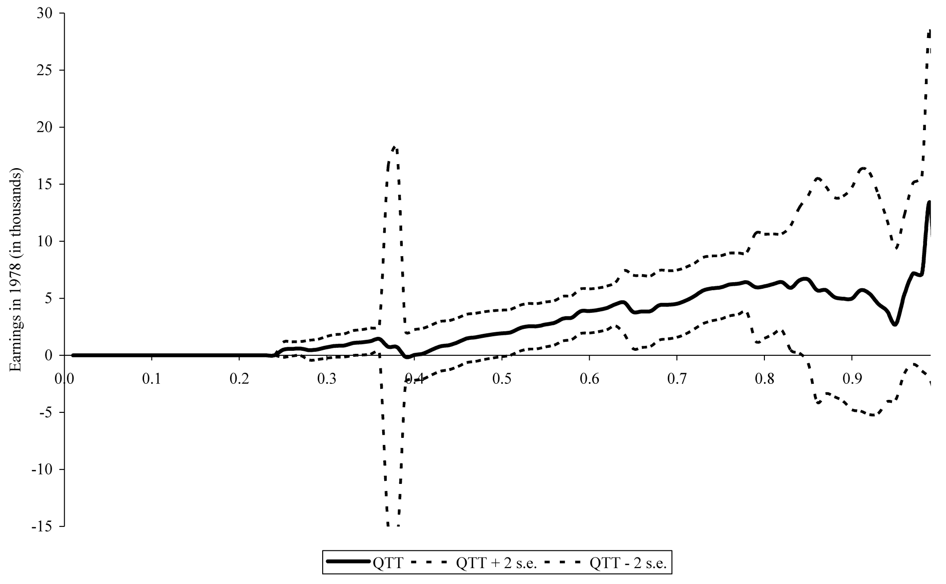

FIGURE 2.—Nonexperimental quantile treatment effects on the treated.

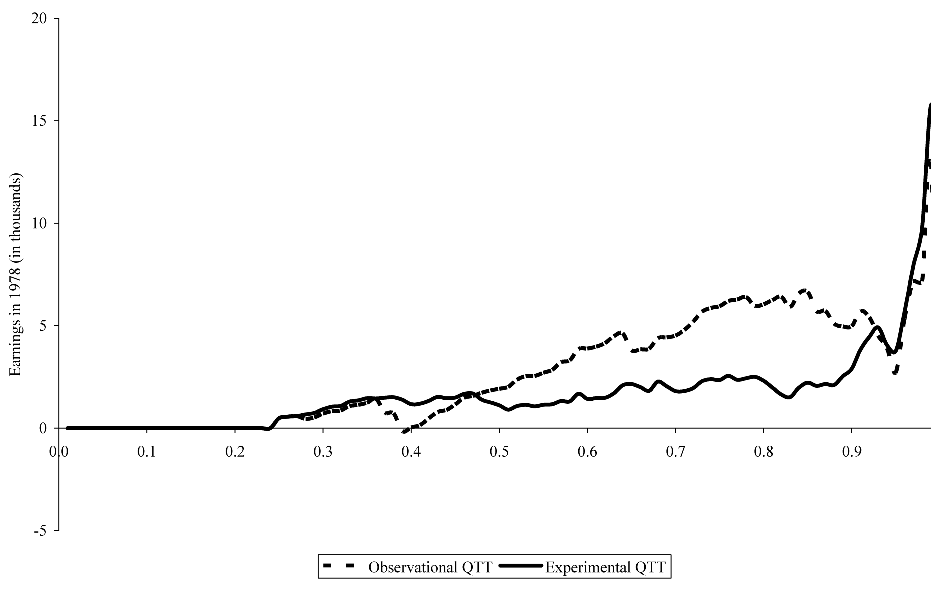

FIGURE 3.—Experimental and nonexperimental QTT.

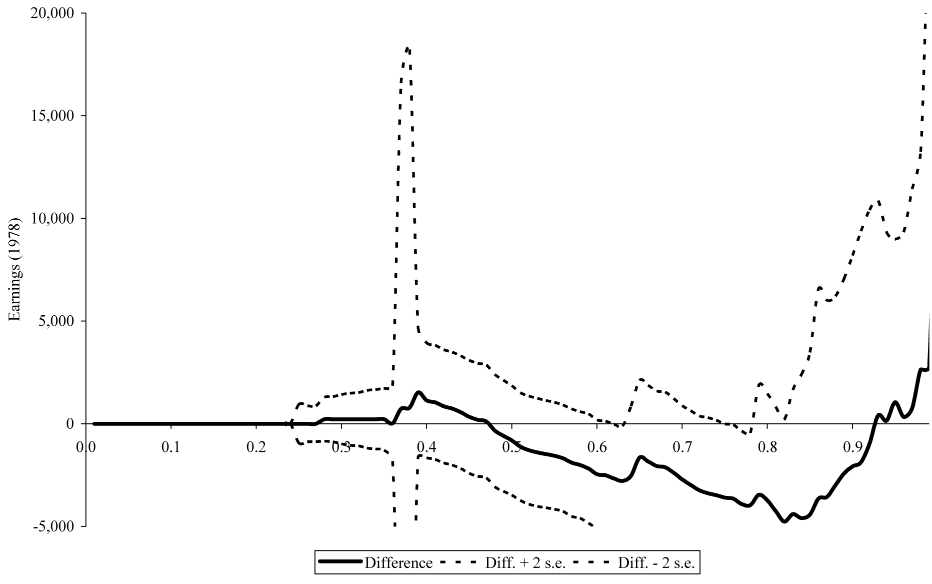

FIGURE 4.—Differences between experimental and nonexperimental QTT.

which instead emphasize selecting propensity-score models that are able to balance each covariate average given propensity-score groups between treated and comparison units.

The chosen model using this nonparametric method includes—in addition to the constant and all linear terms—squared age, squared education, and

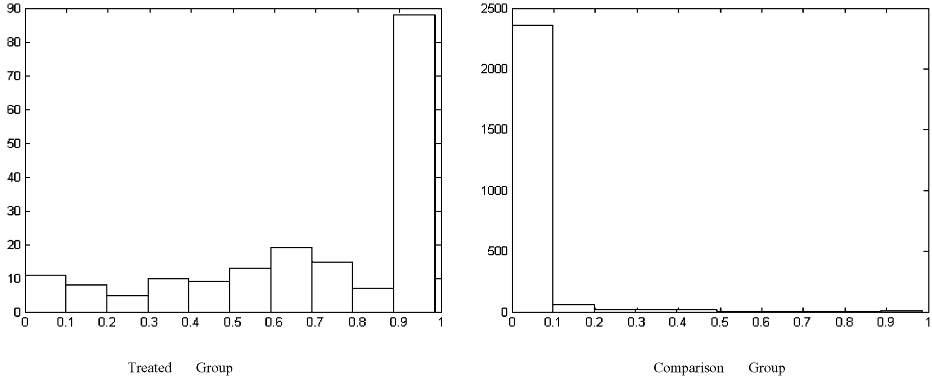

FIGURE 5.—Histograms of estimated propensity score by treatment status.

squared earnings in 1974.[4] Note that despite the difference with the Dehejia and Wahba method to estimate the propensity score, our estimates for average treatment effect on the treated (ATT) and QTT do not vary much if we use their specification instead. For example, using the Dehejia and Wahba (1999) specification for the propensity score, ATT = 1,120 and QTT(0.5) = 1,927, whereas with our specification, ATT = 1,163, but still QTT(0.5) = 1,927.

Estimation of the asymptotic standard errors was performed and, to simplify the estimation procedure, we did not use a cross-validation method in this step, but used the same polynomial selected from the propensity-score estimation. Also, as discussed earlier, density estimation of potential outcomes is necessary for proper standard errors calculations. For the logarithm of positive earnings, we used a Gaussian kernel with the optimal bandwidth (optimal for the normal random variable case).

In Figure 1 we could observe that the counterfactual distribution presents some discrete jumps in the cumulative distribution function (c.d.f.), as if some points had probability mass. A closer look at the data reveals that these points are affected by nonexperimental comparison observations that have the largest values of the estimated propensity score and, therefore, the largest weights. For example, in Figure 1, the percentiles 39–63 of the counterfactual distribution are exactly \$2,305, that is, that point concentrates more than 20% of the total counterfactual distribution. There are, however, only two individuals in the nonexperimental comparison group who reported earnings of \$2,305. One of those two individuals had an estimated propensity score larger than 0.98, which leads to a weight (0.36) that is much larger than the median weight (10^{-6}) of that comparison group.

It is, therefore, no surprise that, with limited overlap, the QTT estimates that are largely influenced by only a few observations have very high standard errors. This can be seen directly in the formula of the asymptotic variance in Equation (10), in particular its second term, which is multiplied by $p(X)/(1 - p(X))$ inside the unconditional expectation bracket. Note that this leads to a very interesting and nonintuitive case: estimators of middle quantiles may have larger standard errors than those of some extreme quantiles because it is possible that, in some cases, the reweighting method will act less intensively in the extreme part of the distribution, offsetting the fact that we have to divide the variance by a very small number, which is the square of the density evaluated at the extreme quantile.

2. A MONTE CARLO STUDY

In this section we report the results of Monte Carlo exercises. The goal is to learn how the estimators of the overall quantile treatment effect (QTE)

[4]Note that this is different from the Dehejia and Wahba (1999) logit specification for the same data set. The difference is that they also include squared 1975 earnings, and an interaction term between the dummy for Hispanic and the dummy for unemployed in 1974.

and the estimator of their asymptotic variance behave in finite samples. The generated data follow a very simple specification. Starting with $X = [X_1, X_2]^\intercal$, we set

$$X_1 \sim \text{Unif}\left[\mu_{X_1} - \frac{\sqrt{12}}{2}, \mu_{X_1} + \frac{\sqrt{12}}{2}\right]$$

and

$$X_2 \sim \text{Unif}\left[\mu_{X_2} - \frac{\sqrt{12}}{2}, \mu_{X_2} + \frac{\sqrt{12}}{2}\right],$$

which will be independent random variables with the following means and variances: $E[X_1] = \mu_{X_1}$, $E[X_2] = \mu_{X_2}$, and $V[X_1] = V[X_2] = 1$. The treatment indicator is set to be $T = \mathbb{1}\{\delta_0 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_1^2 + \eta > 0\}$, where η has a standard logistic c.d.f. $F_\eta(n) = (1 + \exp(-\pi n/\sqrt{3}))^{-1}$. The potential outcomes are $Y(0) = \gamma_1 X_1 + \gamma_2 X_2 + \epsilon_0$ and $Y(1) = Y(0) + \epsilon_1 - \epsilon_0$, where ϵ_0 and ϵ_1 are, respectively, distributed as $N(0, \sigma_{\epsilon_0}^2)$ and $N(\beta, \sigma_{\epsilon_1}^2)$. The variables X, η, ϵ_0, and ϵ_1 are mutually independent. Under this specification, $Y(1)$ and $Y(0)$ will be distributed as the sum of two uniforms and a normal. Their exact distributions for $Y(0)$ and $Y(1)$ are, respectively,

$$\begin{aligned}
&\Pr[Y(0) \le y] \\
&\quad = \frac{\sigma_{\epsilon_0}}{(\overline{z}_1 - \underline{z}_1) \cdot (\overline{z}_2 - \underline{z}_2)} \\
&\qquad \times \int_{-\infty}^{y} \left(\phi\left(\frac{z - (\overline{z}_1 + \overline{z}_2)}{\sigma_{\epsilon_0}}\right) - \phi\left(\frac{z - (\overline{z}_1 + \underline{z}_2)}{\sigma_{\epsilon_0}}\right) \right. \\
&\qquad\qquad - \left(\phi\left(\frac{z - (\underline{z}_1 + \overline{z}_2)}{\sigma_{\epsilon_0}}\right) - \phi\left(\frac{z - (\underline{z}_1 + \underline{z}_2)}{\sigma_{\epsilon_0}}\right) \right) \\
&\qquad\qquad + \Phi\left(\frac{z - (\overline{z}_1 + \overline{z}_2)}{\sigma_{\epsilon_0}}\right) \cdot \left(\frac{z - (\overline{z}_1 + \overline{z}_2)}{\sigma_{\epsilon_0}}\right) \\
&\qquad\qquad - \Phi\left(\frac{z - (\overline{z}_1 + \underline{z}_2)}{\sigma_{\epsilon_0}}\right) \cdot \left(\frac{z - (\overline{z}_1 + \underline{z}_2)}{\sigma_{\epsilon_0}}\right) \\
&\qquad\qquad - \left(\Phi\left(\frac{z - (\underline{z}_1 + \overline{z}_2)}{\sigma_{\epsilon_0}}\right) \cdot \left(\frac{z - (\underline{z}_1 + \overline{z}_2)}{\sigma_{\epsilon_0}}\right) \right. \\
&\qquad\qquad \left.\left. - \Phi\left(\frac{z - (\underline{z}_1 + \underline{z}_2)}{\sigma_{\epsilon_0}}\right) \cdot \left(\frac{z - (\underline{z}_1 + \underline{z}_2)}{\sigma_{\epsilon_0}}\right) \right) \right) dz
\end{aligned}$$

and

$$
\begin{aligned}
&\Pr[Y(1) \le y] \\
&= \frac{\sigma_{\epsilon_1}}{(\overline{z}_1 - \underline{z}_1) \cdot (\overline{z}_2 - \underline{z}_2)} \\
&\quad \times \int_{-\infty}^{y} \Bigg(\phi\left(\frac{z - (\overline{z}_1 + \overline{z}_2 + \beta)}{\sigma_{\epsilon_1}}\right) - \phi\left(\frac{z - (\overline{z}_1 + \underline{z}_2 + \beta)}{\sigma_{\epsilon_1}}\right) \\
&\qquad - \left(\phi\left(\frac{z - (\underline{z}_1 + \overline{z}_2 + \beta)}{\sigma_{\epsilon_1}}\right) - \phi\left(\frac{z - (\underline{z}_1 + \underline{z}_2 + \beta)}{\sigma_{\epsilon_1}}\right)\right) \\
&\qquad + \Phi\left(\frac{z - (\overline{z}_1 + \overline{z}_2 + \beta)}{\sigma_{\epsilon_1}}\right) \cdot \left(\frac{z - (\overline{z}_1 + \overline{z}_2 + \beta)}{\sigma_{\epsilon_1}}\right) \\
&\qquad - \Phi\left(\frac{z - (\overline{z}_1 + \underline{z}_2 + \beta)}{\sigma_{\epsilon_1}}\right) \cdot \left(\frac{z - (\overline{z}_1 + \underline{z}_2 + \beta)}{\sigma_{\epsilon_1}}\right) \\
&\qquad - \Bigg(\Phi\left(\frac{z - (\underline{z}_1 + \overline{z}_2 + \beta)}{\sigma_{\epsilon_1}}\right) \cdot \left(\frac{z - (\underline{z}_1 + \overline{z}_2 + \beta)}{\sigma_{\epsilon_1}}\right) \\
&\qquad - \Phi\left(\frac{z - (\underline{z}_1 + \underline{z}_2 + \beta)}{\sigma_{\epsilon_1}}\right) \cdot \left(\frac{z - (\underline{z}_1 + \underline{z}_2 + \beta)}{\sigma_{\epsilon_1}}\right)\Bigg)\Bigg)\, dz,
\end{aligned}
$$

where $\overline{z}_1 = \gamma_1 \cdot (\mu_{X_1} + \frac{\sqrt{12}}{2})$, $\underline{z}_1 = \gamma_1 \cdot (\mu_{X_1} - \frac{\sqrt{12}}{2})$, $\overline{z}_2 = \gamma_2 \cdot (\mu_{X_2} + \frac{\sqrt{12}}{2})$, and $\underline{z}_2 = \gamma_2 \cdot (\mu_{X_2} - \frac{\sqrt{12}}{2})$; $\Phi(\cdot)$ and $\phi(\cdot)$ are, respectively, the c.d.f. and the probability density function of the standard normal.

The parameters were chosen to be $\mu_{X_1} = 1$, $\mu_{X_2} = 5$, $\delta_0 = -1$, $\delta_1 = 5$, $\delta_2 = -5$, $\delta_3 = -0.05$, $\gamma_1 = -5$, $\gamma_2 = 1$, $\beta = 5$, $\sigma_{\epsilon_0} = 5$, and $\sigma_{\epsilon_1} = 2.5$. Under that specification for the parameters, $Y(1)$ and $Y(0)$ will have, respectively, means and variances equal to $E[Y(1)] = 5$, $E[Y(0)] = 0$, $V[Y(1)] = 56.25$, and $V[Y(0)] = 75$. Also, the population overall quantile treatment effect at, for example, percentiles 0.10, 0.25, 0.50, 0.75, and 0.90 will be, respectively $\Delta_{0.10} = -2.53 - (-9.29) = 6.76$, $\Delta_{0.25} = 0.61 - (-5.04) = 5.65$, $\Delta_{0.50} = 5 - 0 = 5$, $\Delta_{0.75} = 9.39 - 5.04 = 4.35$, and $\Delta_{0.90} = 12.53 - 9.29 = 3.24$.

One thousand replications of this experiment with sample sizes of 500 and 5,000 were considered. Whereas $Y(1)$ and $Y(0)$ are known for each observation i, we could also compute "unfeasible" estimators of parameters of the marginal distributions of $Y(1)$ and $Y(0)$. Because the specification leads to selection on observables, estimation of treatment effects that do not take the selection into account will inevitably produce inconsistent estimates. Those estimates are reported under the label "naive" estimators.

We have calculated average treatment effects and quantile treatment effects. Results can be found in Tables III–VI. Analytical standard errors are also computed.

TABLE III

POINT ESTIMATES FROM MONTE CARLO EXERCISE (SAMPLE SIZE 500; REPLICATIONS 1,000)

Treatment Effects	Estimates		Average	Median	Lower 5th Percentile	Upper 5th Percentile	Standard Deviation	Bias	RMSE	MAE	90% Coverage Rate (%)
Mean	Target	5.0									
	Unfeasible		5.0	5.0	4.6	5.4	0.2	0.0	0.2	0.2	90
	Feasible		5.0	5.0	4.4	5.6	0.4	0.0	0.4	0.3	91
	Naive		3.4	3.4	2.5	4.5	0.6	−1.6	1.7	1.2	17
10th percentile	Target	6.8									
	Unfeasible		6.8	6.8	5.9	7.6	0.5	0.0	0.5	0.4	91
	Feasible		6.8	6.8	5.5	8.1	0.8	0.0	0.8	0.5	90
	Naive		5.6	5.6	4.3	7.0	0.9	−1.1	1.4	1.1	62
25th percentile	Target	5.6									
	Unfeasible		5.7	5.6	5.0	6.4	0.4	0.0	0.4	0.3	90
	Feasible		5.7	5.7	4.6	6.8	0.7	0.1	0.7	0.4	89
	Naive		4.2	4.1	2.8	5.5	0.8	−1.5	1.7	1.5	41
Median	Target	5.0									
	Unfeasible		5.0	5.0	4.4	5.6	0.4	0.0	0.4	0.3	91
	Feasible		5.0	5.0	3.9	6.1	0.7	0.0	0.7	0.4	90
	Naive		3.1	3.1	1.6	4.5	0.9	−1.9	2.1	1.9	29
75th percentile	Target	4.4									
	Unfeasible		4.4	4.4	3.7	5.1	0.4	0.0	0.4	0.3	91
	Feasible		4.3	4.3	3.1	5.6	0.7	0.0	0.7	0.5	89
	Naive		2.6	2.5	1.0	4.2	1.0	−1.8	2.0	1.8	41
90th percentile	Target	3.2									
	Unfeasible		3.2	3.3	2.3	4.2	0.6	0.0	0.6	0.4	90
	Feasible		3.2	3.2	1.6	4.7	0.9	−0.1	0.9	0.6	91
	Naive		1.8	1.8	0.1	3.5	1.0	−1.4	1.8	1.4	61

TABLE IV

POINT ESTIMATES FROM MONTE CARLO EXERCISE (SAMPLE SIZE 5,000; REPLICATIONS 1,000)

Treatment Effects	Estimates		Average	Median	Lower 5th Percentile	Upper 5th Percentile	Standard Deviation	Bias	RMSE	MAE	90% Coverage Rate (%)
Mean	Target	??									
	Unfeasible		5.0	5.0	4.9	5.1	0.1	0.0	0.1	0.1	91
	Feasible		5.0	5.0	4.8	5.2	0.1	0.0	0.1	0.1	90
	Naive		3.4	3.4	3.1	3.7	0.2	–1.6	1.6	1.2	0
10th percentile	Target	6.8									
	Unfeasible		6.8	6.8	6.5	7.1	0.2	0.0	0.2	0.1	90
	Feasible		6.8	6.8	6.4	7.2	0.2	0.0	0.3	0.2	90
	Naive		5.6	5.6	5.2	6.1	0.3	–1.2	1.2	1.2	1
25th percentile	Target	5.6									
	Unfeasible		5.6	5.6	5.4	5.9	0.1	0.0	0.1	0.1	89
	Feasible		5.7	5.7	5.3	6.0	0.2	0.0	0.2	0.1	90
	Naive		4.1	4.1	3.7	4.5	0.3	–1.5	1.6	1.5	0
Median	Target	5.0									
	Unfeasible		5.0	5.0	4.8	5.2	0.1	0.0	0.1	0.1	90
	Feasible		5.0	5.0	4.7	5.3	0.2	0.0	0.2	0.1	90
	Naive		3.0	3.0	2.6	3.5	0.3	–2.0	2.0	2.0	0
75th percentile	Target	4.4									
	Unfeasible		4.4	4.3	4.1	4.6	0.1	0.0	0.1	0.1	91
	Feasible		4.3	4.3	4.0	4.7	0.2	0.0	0.2	0.2	89
	Naive		2.6	2.6	2.1	3.1	0.3	–1.8	1.8	1.8	0
90th percentile	Target	3.2									
	Unfeasible		3.2	3.2	3.0	3.5	0.2	0.0	0.2	0.1	89
	Feasible		3.2	3.2	2.7	3.7	0.3	0.0	0.3	0.2	90
	Naive		1.9	1.9	1.4	2.4	0.3	–1.3	1.4	1.3	1

TABLE V

ANALYTICAL STANDARD ERRORS FROM MONTE CARLO EXERCISE
(SAMPLE SIZE 500; REPLICATIONS 1,000)

Treatment Effects	Estimates	Average	Median	Lower 5th Percentile	Upper 5th Percentile
Mean	Unfeasible	0.2	0.2	0.2	0.3
	Feasible	0.4	0.4	0.3	0.5
	Naive	0.6	0.6	0.6	0.7
10th percentile	Unfeasible	0.5	0.5	0.4	0.6
	Feasible	0.9	0.9	0.7	1.0
	Naive	0.9	0.9	0.8	1.1
25th percentile	Unfeasible	0.5	0.5	0.4	0.5
	Feasible	0.8	0.8	0.7	0.9
	Naive	0.9	0.9	0.8	1.0
Median	Unfeasible	0.4	0.4	0.4	0.5
	Feasible	0.8	0.8	0.7	0.9
	Naive	0.9	0.9	0.8	1.0
75th percentile	Unfeasible	0.5	0.5	0.4	0.5
	Feasible	0.9	0.8	0.7	1.0
	Naive	1.0	0.9	0.8	1.1
90th percentile	Unfeasible	0.5	0.5	0.4	0.6
	Feasible	1.0	1.0	0.8	1.3
	Naive	1.0	1.0	0.9	1.2

The polynomial order for the propensity-score estimation was determined by cross-validation. To simplify our computations, we used the same specification for the variance estimation. Density estimation used a Gaussian kernel with optimal bandwidth.[5]

The results indicate that the reweighting estimator performs well according to both RMSE (root mean squared error) and MAE (median absolute error) criteria. Also, looking separately at bias and variance terms, it is clear that the bias vanishes relatively fast as the sample increases for all of the parameters being estimated by the reweighting method, whereas the same result does not occur with the "naive" estimator, which is an estimator of simple differences between treated and control groups. Analytical standard errors tend to be (either looking at the average or at the median) close to the bootstrapped standard errors for all sample sizes and all parameters. This indicates that, bootstrapping may be a good alternative to analytical standard errors estimation.

[5]See the discussion in Section 3.

TABLE VI

ANALYTICAL STANDARD ERRORS FROM MONTE CARLO EXERCISE
(SAMPLE SIZE 5,000; REPLICATIONS 1,000)

Treatment Effects	Estimates	Average	Median	Lower 5th Percentile	Upper 5th Percentile
Mean	Unfeasible	0.1	0.1	0.1	0.1
	Feasible	0.1	0.1	0.1	0.1
	Naive	0.2	0.2	0.2	0.2
10th percentile	Unfeasible	0.2	0.2	0.2	0.2
	Feasible	0.3	0.3	0.3	0.3
	Naive	0.3	0.3	0.3	0.3
25th percentile	Unfeasible	0.1	0.1	0.1	0.1
	Feasible	0.2	0.2	0.2	0.3
	Naive	0.3	0.3	0.3	0.3
Median	Unfeasible	0.1	0.1	0.1	0.1
	Feasible	0.2	0.2	0.2	0.3
	Naive	0.3	0.3	0.3	0.3
75th percentile	Unfeasible	0.1	0.1	0.1	0.1
	Feasible	0.3	0.3	0.2	0.3
	Naive	0.3	0.3	0.3	0.3
90th percentile	Unfeasible	0.2	0.2	0.2	0.2
	Feasible	0.3	0.3	0.3	0.3
	Naive	0.3	0.3	0.3	0.3

The results constitute strong evidence that, even for relatively small sample sizes (e.g., 500), the estimator would have good precision and that inference based on a first-order asymptotic approximation would be appropriate.

3. DETAILS OF VARIANCE ESTIMATION PROCEDURE

The proposed estimator for V_τ is $\widehat{V}_\tau = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\varphi}_{\tau,i} + \widehat{\alpha}_{\tau,i})^2$, where

$$\widehat{\varphi}_{\tau,i} = \frac{T_i}{\widehat{p}(X_i)} \cdot \widehat{g}_{1,\tau}(Y_i) - \frac{1-T_i}{1-\widehat{p}(X_i)} \cdot \widehat{g}_{0,\tau}(Y_i),$$

$$\widehat{\alpha}_{\tau,i} = -\widehat{E}\left[\frac{T \cdot \widehat{g}_{1,\tau}(Y)}{(\widehat{p}(X))^2} + \frac{(1-T)\cdot \widehat{g}_{0,\tau}(Y)}{(1-\widehat{p}(X))^2}\middle| X = X_i\right] \cdot (T_i - \widehat{p}(X_i)),$$

$$\widehat{g}_{j,\tau}(Y) = -\frac{\mathbb{1}\{Y \le \widehat{q}_{j,\tau}\} - \tau}{\widehat{f}_j(\widehat{q}_{j,\tau})}, \qquad (j = 0, 1).$$

There are several parts of this estimation process that need to be specified. Consider first $\widehat{f}_j(\cdot)$, which is an estimator of the density of the potential outcome $Y(j)$,

$$\widehat{f}_j(y) = \sum_{i=1}^{N} \widehat{\omega}_{j,i} \cdot \mathcal{K}_{h_j,j}(Y_i - y),$$

where $\mathcal{K}_{h_j,j}(\cdot) = h_j^{-1}\mathcal{K}_j(\cdot/h_j)$, and where $\mathcal{K}_j(\cdot)$ is a kernel function and h_j is a bandwidth.[6] This is the reweighted kernel estimator method, proposed in DiNardo, Fortin, and Lemieux (1996), that is used because simple kernel estimators are unable to address the missing data problem present in the estimation of the marginal density of the potential outcomes.

Next, consider $\widehat{\alpha}_{\tau,i}$, which is an estimator of an unknown conditional expectation times $(\widehat{p}(X_i) - T_i)$. Following Hirano, Imbens, and Ridder (2003, henceforth HIR), we propose estimating that conditional expectation by a polynomial approximation, which is exactly the same estimation procedure used for the log-odds ratio of the propensity score, but possibly using a different vector H_{K_τ} of polynomials of x for $\tau \in (0, 1)$.[7,8] Basically:

$$\widehat{E}\left[\frac{T \cdot \widehat{g}_{1,\tau}(Y)}{(\widehat{p}(X))^2} + \frac{(1-T) \cdot \widehat{g}_{0,\tau}(Y)}{(1-\widehat{p}(X))^2}\middle| X = X_i\right] = H_{K_\tau}(X_i)' \cdot \widehat{\lambda}_\tau,$$

where

$$\widehat{\lambda}_\tau = \left(\sum_{l=1}^{N} H_{K_\tau}(X_l) \cdot H_{K_\tau}(X_l)'\right)^{-1} \times \left(\sum_{i=1}^{N} H_{K_\tau}(X_i) \cdot \left(\frac{T_i \cdot \widehat{g}_{1,\tau}(Y_i)}{(\widehat{p}(X_i))^2} + \frac{(1-T_i) \cdot \widehat{g}_{0,\tau}(Y_i)}{(1-\widehat{p}(X_i))^2}\right)\right).$$

4. PROOFS

PROOF OF LEMMA 1: Starting from the definition of the τ quantile of $Y(1)$, we show how to express $q_{1,\tau}$ in terms of the observed data (Y, T, X):

[6] In the Monte Carlo exercise as well as in the empirical application, we have used the standard normal kernel with optimal bandwidth for that distribution.

[7] We propose using vectors of polynomials with the same properties as those used for the propensity-score estimation. That is, for $\tau \in (0, 1)$, let $H_{K_\tau}(x) = [H_{K_\tau,l}(x)]$, $l = 1, \ldots, K_\tau$, be a vector of length K_τ of polynomial functions of $x \in \mathcal{X}$ that satisfy the properties (i) $H_{K_\tau} : \mathcal{X} \to \mathbb{R}^{K_\tau}$ and (ii) $H_{K_\tau,1}(x) = 1$. Also, $K_\tau = K_\tau(N) \to \infty$ as $N \to \infty$ and (iii) if $K_\tau > (n+1)^r$, then $H_{K_\tau}(x)$ includes all polynomials up to order n.

[8] For computational simplicity in the Monte Carlo exercise as well as in the empirical application, we have fixed the vectors of polynomial functions H_{K_τ} to be the same as in the propensity-score estimation.

$\tau = \Pr[Y(1) \leq q_{1,\tau}] = E[\Pr[Y(1) \leq q_{1,\tau}|X]] = E[\Pr[Y(1) \leq q_{1,\tau}|X, T = 1]] = E[\Pr[Y \leq q_{1,\tau}|X, T = 1]] = E[E[T\mathbb{1}\{Y \leq q_{1,\tau}\}|X, T = 1]] = E[\frac{1}{p(X)} E[T\mathbb{1}\{Y \leq q_{1,\tau}\}|X]] = E[\frac{T}{p(X)}\mathbb{1}\{Y \leq q_{1,\tau}\}]$.

The first equality follows from the definition of $q_{1,\tau}$ and from Assumption 2. The second equality is an application of the law of iterated expectations. The third equality follows from the ignorability assumption (Assumption 1). The fourth equality results from the definition of $Y = TY(1) + (1 - T)Y(0)$. The fifth equality comes from $E[\mathbb{1}\{A\}] = \Pr[A]$ (where A is some event) and from the fact that the expectation is conditional on $T = 1$. The sixth equality is a consequence from $E[Z|X] = p(X)E[Z|X, T = 1] + (1 - p(X))E[Z|X, T = 0]$, where Z is some random variable. Finally, the last equality is a backward application of the law of iterated expectations. Analogous results for $q_{0,\tau}$, $q_{1,\tau|T=1}$, and $q_{0,\tau|T=1}$ could have been derived following essentially the same steps. *Q.E.D.*

PROOF OF COROLLARY 1: Note that, from Lemma 1, the four parameters $q_{1,\tau}$, $q_{0,\tau}$, $q_{1,\tau|T=1}$, and $q_{0,\tau|T=1}$ are functionals of the joint distribution of (Y, T, X). Whereas Δ_τ equals the difference between $q_{1,\tau}$ and $q_{0,\tau}$, and $\Delta_{\tau|T=1}$ equals the difference between $q_{1,\tau|T=1}$ and $q_{0,\tau|T=1}$, Δ_τ and $\Delta_{\tau|T=1}$ are also functionals of the joint distribution of (Y, T, X). Therefore, Δ_τ and $\Delta_{\tau|T=1}$, are identified from data on (Y, T, X). *Q.E.D.*

PROOF OF THEOREM 1: The proof is divided into five parts:

Part 1: *Asymptotic Properties of the First Step*. The approach described in the main article to estimating the propensity score guarantees, under certain regularity conditions, that $\widehat{p}(x)$, the estimator of the propensity score, is uniformly consistent for the true $p(x)$. To assure that this holds, we make the following assumptions:

ASSUMPTION A.1—First Step:
(i) *The parameter* $\mathcal{X}$ *is a compact subset of* $\mathbb{R}^r$.
(ii) *The density of* X, $f(x)$, *satisfies* $0 < \inf_{x\in\mathcal{X}} f(x) \leq \sup_{x\in\mathcal{X}} f(x) < \infty$.
(iii) *The polynomial* $p(x)$ *is s-times continuously differentiable*, *where* $s \geq 7r$ *and* r *is the dimension of* X.
(iv) *The order of* $H_K(x)$, K, *is of the form* $K = C \cdot N^c$, *where* C *is a constant and* $c \in (\frac{1}{4(s/r-1)}, \frac{1}{9})$.

Newey (1995, 1997) established that, for orthogonal polynomials $H_K(x)$ and compact $\mathcal{X}$,

$$\Gamma(K) = \sup_{x\in\mathcal{X}} \|H_K(x)\| \leq C \cdot K, \tag{A-1}$$

where C is a generic constant. Note then that because of part (iv) of Assumption A.1, Γ will be a function of N because K is assumed to be a function of N.

With part (ii) of Assumption 1 (Common Support) and Assumption A.1 in hand, we can invoke some of the results derived by HIR in the format of a lemma:

LEMMA A.1—First Step: *Under Assumptions* 1 *and* A.1, *the following results hold*:

(I) *There exists* $\sup_{x\in\mathcal{X}}|p(x)-p_K(x)|\leq C\cdot\Gamma(K)\cdot K^{-s/2r}\leq C\cdot K^{1-s/2r}\leq C\cdot N^{(1-s/2r)c}=o(1)$, *where* $p_K(x)=L(H_K(x)'\pi_K)$ *and* $\pi_K=\arg\max_{\pi\in\mathbb{R}^K}E\{p(X)\cdot\log(L(H_K(X)'\pi))+(1-p(X))\cdot\log(1-L(H_K(X)'\pi))\}$.

(II) *There exists* $\|\widehat{\pi}_K-\pi_K\|=O_p(\sqrt{K(N)/N})\leq C\cdot O_p(\sqrt{N^c/N})\leq C\cdot O_p(N^{(c-1)/2})=o_p(1)$.

(III) *There exists* $\sup_{x\in\mathcal{X}}|p(x)-\widehat{p}(x)|\leq C_1\cdot N^{(1-s/2r)c}+O_p(\Gamma(K)\cdot\sqrt{K(N)/N})\leq C_1\cdot N^{(1-s/2r)c}+C_2\cdot N^{(3c-1)/2}\cdot O_p(1)=o_p(1)$.

(IV) *There exists* $\varepsilon>0$, $\lim_{N\to\infty}\Pr[\varepsilon<\inf_{X\in\mathcal{X}}\widehat{p}(X)\leq\sup_{X\in\mathcal{X}}\widehat{p}(X)<1-\varepsilon]=1$.

See Hirano, Imbens, and Ridder (2003) for the proof.

Part 2: *Change of Variables—u and Q_N*. Recalling from the minimization problem described by Equation (4) that

$$\begin{aligned}\widehat{q}_{1,\tau}&=\arg\min_q G_{\tau,N}(q,\widehat{p})\\&=\arg\min_q\frac{1}{N}\sum_{i=1}^N\frac{T_i}{\widehat{p}(X_i)}(Y_i-q)(\tau-\mathbb{1}\{Y_i\leq q\}),\end{aligned}$$

we have then

$$\begin{aligned}\widehat{q}_{1,\tau}&=\arg\min_q N\cdot\big(G_{\tau,N}(q,\widehat{p})-G_{\tau,N}(q_{1,\tau},\widehat{p})\big)\\&=\arg\min_q\sum_{i=1}^N\frac{T_i}{\widehat{p}(X_i)}\big[(Y_i-q)(\tau-\mathbb{1}\{Y_i\leq q\})\\&\qquad\qquad-(Y_i-q_{1,\tau})(\tau-\mathbb{1}\{Y_i\leq q_{1,\tau}\})\big]\\&=\arg\min_q\sum_{i=1}^N\frac{T_i}{\widehat{p}(X_i)}\big[(\mathbb{1}\{Y_i\leq q_{1,\tau}\}-\tau)(q-q_{1,\tau})\\&\qquad\qquad+(Y_i-q)(\mathbb{1}\{Y_i\leq q_{1,\tau}\}-\mathbb{1}\{Y_i\leq q\})\big].\end{aligned}$$

Now, we define: $u \equiv \sqrt{N}(q - q_{1,\tau})$, $\widehat{u}_\tau \equiv \sqrt{N}(\widehat{q}_{1,\tau} - q_{1,\tau})$,

$$\text{(A-2)} \qquad \begin{aligned} D_\tau(Y_i) &\equiv \mathbb{1}\{Y_i \le q_{1,\tau}\} - \tau, \\ R_\tau(Y_i, u) &\equiv \big(Y_i - (q_{1,\tau} + u/\sqrt{N}\,)\big) \\ &\qquad \times \big(\mathbb{1}\{Y_i \le q_{1,\tau}\} - \mathbb{1}\{Y_i \le q_{1,\tau} + u/\sqrt{N}\,\}\big), \\ Q_{\tau,N}(u, \widehat{p}) &\equiv \sum_{i=1}^{N} \frac{T_i}{\widehat{p}(X_i)} \cdot \left(D_\tau(Y_i) \cdot \frac{u}{\sqrt{N}} + R_\tau(Y_i, u) \right) \\ &= N \cdot \big(G_{\tau,N}(q, \widehat{p}) - G_{\tau,N}(q_{1,\tau}, \widehat{p})\big). \end{aligned}$$

A comment about some of the preceding quantities: The variable $D_\tau(Y_i)$ is the approximate first derivative of the check function $\rho_\tau(Y_i - q)$ with respect to q. It is approximate in the sense that $\rho_\tau(Y_i - q)$ is not differentiable for all q, because it involves indicator functions of whether q is less than or equal to some values in the data. The variable $R_\tau(Y_i, u)$ can be interpreted as the remainder term from a linear expansion about $q_{1,\tau}$ that uses $D_\tau(Y_i)$ as an approximated derivative.

Next, note that whereas $\widehat{u}_\tau = \sqrt{N}(\widehat{q}_{1,\tau} - q_{1,\tau})$, then, by Equation (4), $\widehat{u}_\tau = \arg\min_u Q_{\tau,N}(u, \widehat{p})$.

Part 3: *A Quadratic Approximation to the Objective Function*. We now show that $\xi_{\tau,N}(u, \widehat{p}) = Q_{\tau,N}(u, \widehat{p}) - \widetilde{Q}_{\tau,N}(u)$ is $o_p(1)$ for fixed u, where $\widetilde{Q}_{\tau,N}(u)$ is a quadratic random function in u that does not depend on $\widehat{p}$, the estimated p-score.

We begin by defining the function

$$\text{(A-3)} \qquad \begin{aligned} \widetilde{Q}_{\tau,N}(u) = u \cdot \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \Bigg(\frac{T_i}{p(X_i)} \cdot \big(D_\tau(Y_i) - E[D_\tau(Y)|X_i, T = 1]\big) \\ + E[D_\tau(Y)|X_i, T = 1] \Bigg) \\ + u^2 \cdot \frac{f_1(q_{1,\tau})}{2} \end{aligned}$$

and by writing the absolute value of the difference as

$$\begin{aligned} &|Q_{\tau,N}(u, \widehat{p}) - \widetilde{Q}_{\tau,N}(u, \widehat{p})| \\ &\quad = \Bigg| \frac{u}{N^{1/2}} \sum_{i=1}^{N} \frac{T_i}{\widehat{p}(X_i)} \cdot D_\tau(Y_i) \end{aligned}$$

$$
\begin{aligned}
&- \left(\frac{T_i}{p(X_i)} \cdot \big(D_\tau(Y_i) - E[D_\tau(Y)|X_i, T=1]\big) \right. \\
&\qquad \left. + E[D_\tau(Y)|X_i, T=1] \right) \\
&+ \sum_{i=1}^{N} \frac{T_i}{\widehat{p}(X_i)} \cdot R_\tau(Y_i, u) - u^2 \cdot \frac{f_1(q_{1,\tau})}{2} \Bigg| \\
\leq\ & |\xi_{1,\tau,N}(u, \widehat{p})| + |\xi_{2,\tau,N}(u, \widehat{p})|,
\end{aligned}
$$

where

$$
\begin{aligned}
\xi_{1,\tau,N}(u, \widehat{p}) = \frac{u}{N^{1/2}} \sum_{i=1}^{N} \frac{T_i}{\widehat{p}(X_i)} \cdot D_\tau(Y_i) \\
- \left(\frac{T_i}{p(X_i)} \cdot \big(D_\tau(Y_i) - E[D_\tau(Y)|X_i, T=1]\big) \right. \\
\left. + E[D_\tau(Y)|X_i, T=1] \right), \\
\xi_{2,\tau,N}(u, \widehat{p}) = \sum_{i=1}^{N} \frac{T_i}{\widehat{p}(X_i)} \cdot R_\tau(Y_i, u) - u^2 \cdot \frac{f_1(q_{1,\tau})}{2}.
\end{aligned}
$$

The next lemma shows that $\xi_{\tau,N}(u, \widehat{p})$ goes to zero in probability for each u, which means that the objective function can be approximated by a quadratic (in u) random function. Before stating the lemma, let us first assume that the next regularity condition holds:

ASSUMPTION A.2—Continuity: *For* $j = 0, 1$, $E[D_\tau(Y(j))|X = x]$ *is continuously differentiable with respect to* x *for every* x *in* $\mathcal{X}$.

LEMMA A.2—Bounding the Differences in the Objective Function: *Under Assumptions* 1, 2, A.1, *and* A.2, *for each* u, $\xi_{\tau,N}(u, \widehat{p}) = o_p(1)$.

PROOF: To prove Lemma A.2, we use the previous decomposition of $\xi_{\tau,N}(u, \widehat{p})$ into the sum of two terms. The first term, $\xi_{1,\tau,N}(u, \widehat{p})$, will be $u \cdot o_p(1)$. This is a direct consequence of Theorem 1 from HIR, applied not to Y as in that paper, but to a bounded function of Y: $D_\tau(Y)$. The requirements for the HIR theorem to hold are verified for our case: (i) Assumptions 1 and A.1 cover Assumptions 1, 2, 4, and 5 of HIR, and (ii) $V[D_\tau(Y(1))] = \tau(1-\tau) < +\infty$ plus Assumption A.2 cover their Assumption 3.

The next part of the sum is $\xi_{2,\tau,N}(u, \widehat{p})$, which can be bounded as

$$|\xi_{2,\tau,N}(u, \widehat{p})| = \left| \sum_{i=1}^{N} \frac{T_i}{\widehat{p}(X_i)} \cdot R_\tau(Y_i, u) - u^2 \cdot \frac{f_1(q_{1,\tau})}{2} \right|$$

$$\leq \left| \sum_{i=1}^{N} T_i \cdot R_\tau(Y_i, u) \cdot \frac{\widehat{p}(X_i) - p(X_i)}{\widehat{p}(X_i) \cdot p(X_i)} \right| \tag{A-4}$$

$$+ \left| \sum_{i=1}^{N} \frac{T_i}{p(X_i)} \cdot R_\tau(Y_i, u) - E\left[\frac{T_i}{p(X_i)} \cdot R_\tau(Y_i, u) \right] \right| \tag{A-5}$$

$$+ \left| N \cdot E\left[\frac{T}{p(X)} \cdot R_\tau(Y, u) \right] - u^2 \cdot \frac{f_1(q_{1,\tau})}{2} \right|. \tag{A-6}$$

The first term of the latter sum (A-4), converges to zero in probability for fixed u, because

$$\left| \sum_{i=1}^{N} \frac{T_i \cdot R_\tau(Y_i, u)}{p(X_i)} \cdot \frac{\widehat{p}(X_i) - p(X_i)}{\widehat{p}(X_i)} \right|$$
$$\leq \sum_{i=1}^{N} \left| \frac{T_i \cdot R_\tau(Y_i, u)}{p(X_i)} \right| \cdot \left(\inf_{x \in \mathcal{X}} \widehat{p}(x) \right)^{-1} \cdot \sup_{x \in \mathcal{X}} (\widehat{p}(x) - p(x))$$
$$= O_p\left(\sqrt{N \cdot E\left[(R_\tau(Y(1), u))^2 \right]} \right) \cdot O_p(1) \cdot o_p(1)$$
$$= O_p\left(N^{1/2} \cdot O\left(\frac{|u|^{3/2}}{N^{3/4}} \right) \right) \cdot o_p(1)$$
$$= O_p(N^{-1/4} \cdot |u|^{3/2}) \cdot o_p(1) = |u|^{3/2} \cdot o_p(1).$$

This result holds because the first and second moments $R_\tau(Y(1), u)$ are given by

$$E\left[R_\tau(Y(1), u) \right]$$
$$= \int_{q_{1,\tau}+u/\sqrt{N}}^{q_{1,\tau}} \left(y - \left(q_{1,\tau} + \frac{u}{\sqrt{N}} \right) \right) f_1(y)\, dy$$
$$= \left(y - \left(q_{1,\tau} + \frac{u}{\sqrt{N}} \right) \right) F_1(y) \Big|_{q_{1,\tau}+u/\sqrt{N}}^{q_{1,\tau}} + \int_{q_{1,\tau}}^{q_{1,\tau}+u/\sqrt{N}} F_1(y)\, dy$$
$$= -F_1(q_{1,\tau}) \frac{u}{\sqrt{N}} + F_1(q_{1,\tau}) \frac{u}{\sqrt{N}} + \frac{u^2}{2N} f_1\left(q_{1,\tau} + \frac{u^*}{\sqrt{N}} \right)$$
$$= \frac{u^2}{2N} \cdot f_1\left(q_{1,\tau} + \frac{u^*}{\sqrt{N}} \right)$$

for some u^* between 0 and u, where $F_1(y) = \Pr[Y(1) \le y]$, and

$$\begin{aligned} &E\big[R_\tau^2(Y(1), u)\big] \\ &= \int_{q_{1,\tau}+u/\sqrt{N}}^{q_{1,\tau}} \left(y - \left(q_{1,\tau} + \frac{u}{\sqrt{N}}\right)\right)^2 \mathbb{1}\{y \le q_{1,\tau}\} f_1(y)\,dy \\ &\quad - \int_{q_{1,\tau}+u/\sqrt{N}}^{q_{1,\tau}} \left(y - \left(q_{1,\tau} + \frac{u}{\sqrt{N}}\right)\right)^2 \mathbb{1}\left\{y \le q_{1,\tau} + \frac{u}{\sqrt{N}}\right\} f_1(y)\,dy. \end{aligned}$$

Consider the case in which $u > 0$:[9]

$$\begin{aligned} &E\big[R_\tau^2(Y(1), u)\big] \\ &= \int_{q_{1,\tau}}^{q_{1,\tau}+u/\sqrt{N}} \left(y - \left(q_{1,\tau} + \frac{u}{\sqrt{N}}\right)\right)^2 f_1(y)\,dy \\ &= \left(y - \left(q_{1,\tau} + \frac{u}{\sqrt{N}}\right)\right)^2 F_1(y)\Bigg|_{q_{1,\tau}}^{q_{1,\tau}+u/\sqrt{N}} \\ &\quad - 2\int_{q_{1,\tau}}^{q_{1,\tau}+u/\sqrt{N}} \left(y - \left(q_{1,\tau} + \frac{u}{\sqrt{N}}\right)\right) F_1(y)\,dy \\ &= -F_1(q_{1,\tau})\frac{u^2}{N} + 2\left(\frac{u^2}{2N}F_1(q_{1,\tau}) + \frac{u^3}{6N^{3/2}} f_1\left(q_{1,\tau} + \frac{u^{**}}{\sqrt{N}}\right)\right) \\ &= \frac{u^3}{3N^{3/2}} f_1\left(q_{1,\tau} + \frac{u^{**}}{\sqrt{N}}\right), \end{aligned}$$

where u^{**} is some real number between 0 and u.

Now, we use similar arguments to find the probability order of the second term (A-5),

$$\begin{aligned} &\left| \sum_{i=1}^{N} \frac{T_i}{p(X_i)} \cdot R_\tau(Y_i, u) - E\left[\frac{T_i}{p(X_i)} \cdot R_\tau(Y_i, u)\right] \right| \\ &= O_p\left(\sqrt{N \cdot V\big[R_\tau(Y(1), u)\big]}\right) \\ &= O_p\left(N^{1/2} \cdot O\left(\frac{|u|^{3/2}}{N^{3/4}}\right)\right) = O_p(N^{-1/4} \cdot |u|^{3/2}) \\ &= o_p(|u|^{3/2}), \end{aligned}$$

[9] The $u < 0$ case yields the same result times (-1).

and finally, the last term (A-6), is

$$\begin{aligned}
&\left| N \cdot E\left[\frac{T}{p(X)} \cdot R_\tau(Y,u)\right] - u^2 \cdot \frac{f_1(q_{1,\tau})}{2}\right| \\
&\quad = N \cdot \left| E\big[R_\tau(Y(1),u)\big] - \frac{u^2}{2N} \cdot f_1(q_{1,\tau})\right| \\
&\quad = N \cdot o\left(\frac{u^2}{N}\right).
\end{aligned}$$

Therefore, $\xi_{2,\tau,N}(u,\widehat{p}) = |u|^{3/2} \cdot o_p(1)$. So, for fixed u, $\xi_{\tau,N}(u,\widehat{p}) = o_p(1)$. *Q.E.D.*

Part 4: *Asymptotic Properties of* $\widetilde{u}_\tau$. We show that $\widetilde{u}_\tau$, the argument that minimizes the random quadratic $\widetilde{Q}_{\tau,N}(u)$, (i) is $O_p(1)$ and (ii) $\widetilde{u}_\tau \xrightarrow{D} N(0,V_{1,\tau})$.

Under Assumption 2, the argument that minimizes $\widetilde{Q}_{\tau,N}(u)$ is unique. Also, using Equation (A-3), it is equal to

$$\begin{aligned}
\text{(A-7)} \qquad \widetilde{u}_\tau &= \arg\min_u \widetilde{Q}_{\tau,N}(u) \\
&= -\frac{1}{\sqrt{N} f_1(q_{1,\tau})} \sum_{i=1}^{N} \left(\frac{T_i}{p(X_i)} \cdot D_\tau(Y_i) \right. \\
&\qquad \left. - \frac{T_i - p(X_i)}{p(X_i)} \cdot E[D_\tau(Y)|X_i, T=1] \right) \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \psi_{1,\tau}(Y_i, T_i, X_i),
\end{aligned}$$

where $\psi_{1,\tau}$ is the influence function of $q_{1,\tau}$, which was defined by Equation (6).

Let us now write the main result of this subsection as a lemma:

LEMMA A.3—Asymptotic Properties of $\widetilde{u}_\tau$: *Let* $\widetilde{u}_\tau = \arg\min_u \widetilde{Q}_{\tau,N}(u)$. *Then, under Assumptions* 1, 2, *and* A.1, $\widetilde{u}_\tau \xrightarrow{D} N(0, E[\psi^2_{1,\tau}(Y,T,X)])$.

PROOF: The result follows by noting that $E[\psi_{1,\tau}(Y,T,X)] = 0$ and by an application of the central limit theorem. *Q.E.D.*

Part 5: *Nearness of arg mins*. We show that the term $\widehat{u}_\tau$ is just $o_p(1)$ from $\widetilde{u}_\tau$ or, written in terms of q, that $\widehat{q}_{1,\tau}$ is asymptotically equivalent to $\widetilde{q}_{1,\tau} = \widetilde{u}_\tau/\sqrt{N} + q_{1,\tau}$.

To get results about $\widehat{u}_\tau$ and consequently about $\widehat{q}_{1,\tau}$, we will use a result in Hjört and Pollard (1993) on the nearness of minimizers of convex random functions. In particular, we apply Hjört and Pollard's Lemma 2 directly to our case:

LEMMA A.4—Nearness of arg mins (Hjört and Pollard (1993)): *Under Assumptions* 1, 2, A.1, *and* A.2, *we have the following probabilistic bound on how far* $\widehat{u}_\tau$ *can be from* $\widetilde{u}_\tau$: *For each* $\varepsilon > 0$,

$$\Pr[|\widehat{u}_\tau - \widetilde{u}_\tau| \geq \varepsilon] \leq \Pr\left[\sup_{|u-\widetilde{u}_\tau|\leq\varepsilon} |\xi_{\tau,N}(u,\widehat{p})| \geq \frac{1}{4} f_1(q_{1,\tau})\varepsilon^2\right]$$

and, *moreover*,

$$\Pr\left[\sup_{|u-\widetilde{u}_\tau|\leq\varepsilon} |\xi_{\tau N}(u,\widehat{p})| \geq \frac{1}{4} f_1(q_{1,\tau})\varepsilon^2\right] = o(1).$$

PROOF: First notice that $G_{\tau,N}(q,\widehat{p}) = (1/N)\sum_{i=1}^N T_i/(\widehat{p}(X_i))\rho_\tau(Y_i - q)$ is convex in q with probability approaching 1, because it is a sum of zeros and convex functions in q. As a result, the transformed objective function $Q_{\tau,N}(u,\widehat{p})$ will be convex in u and the following random function must be convex in u:

$$\begin{aligned} B_{\tau,N}(u,\widehat{p}) &= Q_{\tau,N}(u,\widehat{p}) \\ &\quad - \sum_{i=1}^N \frac{u}{\sqrt{N}}\left(\frac{T_i D_\tau(Y_i)}{p(X_i)} - E[D_\tau|X_i, T=1]\cdot\frac{T_i - p(X_i)}{p(X_i)}\right) \\ &= \frac{1}{2} f_1(q_{1,\tau})u^2 + \xi_{\tau,N}(u,\widehat{p}). \end{aligned} \tag{A-8}$$

Let us call $B_\tau(u)$ the quadratic $\frac{1}{2}f_1(q_{1,\tau})\cdot u^2$. Now, by convexity of $B_{\tau,N}(u,\widehat{p})$ for any u such that $|u-\widetilde{u}_\tau| = a > \varepsilon$,

$$\left(1-\frac{\varepsilon}{a}\right)\cdot B_{\tau,N}(\widetilde{u}_\tau,\widehat{p}) + \frac{\varepsilon}{a}\cdot B_{\tau,N}(u,\widehat{p}) \geq B_{\tau,N}(\widetilde{u}_\tau+\varepsilon,\widehat{p}).$$

By Equation (A-8), this can be rewritten as

$$\begin{aligned} &\frac{\varepsilon}{a}\big(B_{\tau,N}(u,\widehat{p}) - B_{\tau N}(\widetilde{u}_\tau,\widehat{p})\big) \\ &\quad \geq B_\tau(\widetilde{u}_\tau+\varepsilon) + \xi_{\tau,N}(\widetilde{u}_\tau+\varepsilon,\widehat{p}) - (B_\tau(\widetilde{u}_\tau) + \xi_{\tau,N}(\widetilde{u}_\tau,\widehat{p})) \\ &\quad \geq -2\sup_{|u-\widetilde{u}_\tau|\leq\varepsilon}|\xi_{\tau,N}(u,\widehat{p})| + \inf_{|u-\widetilde{u}_\tau|=\varepsilon}|B_\tau(u) - B_\tau(\widetilde{u})|. \end{aligned}$$

Now note that $\inf_{|u-\widetilde{u}_\tau|=\varepsilon}|B_\tau(u) - B_\tau(\widetilde{u}_\tau)| = \frac{1}{2}f_1(q_{1,\tau})\varepsilon^2$. Thus, for all u outside the ε interval around $\widetilde{u}_\tau$, if

$$-2\sup_{|u-\widetilde{u}_\tau|\leq\varepsilon}|\xi_{\tau,N}(u,\widehat{p})| + \frac{1}{2}f_1(q_{1,\tau})\varepsilon^2 > 0, \tag{A-9}$$

then $\widehat{u}_\tau$, the minimizer of $Q_{\tau,N}(u, \widehat{p})$, will be inside the ε interval around $\widetilde{u}_\tau$. Hence, we need to show that, with probability approaching 1, Equation (A-9) holds.

By the Hjört and Pollard's (1993) version of the convexity lemma, $\sup_{u\in\mathcal{C}} |\xi_{\tau,N}(u, \widehat{p})| = o_p(1)$ for each compact subset $\mathcal{C}$ of $\mathbb{R}$. Define $\mathcal{C}_{\tau,\varepsilon} = \{u \in \mathbb{R}; |u - \widetilde{u}_\tau| \leq \varepsilon\}$. Because $\mathcal{C}_{\tau,\varepsilon}$ is a bounded and closed subset of $\mathbb{R}$, it is compact. Therefore, $\sup_{u\in\mathcal{C}_{\tau,\varepsilon}} |\xi_{\tau,N}(u, \widehat{p})| = o_p(1)$. Thus, for each $\varepsilon > 0$, $\Pr[\sup_{u\in\mathcal{C}_{\tau,\varepsilon}} |\xi_{\tau,N}(u, \widehat{p})| \geq \frac{1}{4} f_1(q_{1,\tau})\varepsilon^2] = o(1)$.

Hence, with probability approaching 1, for each $\varepsilon > 0$, Equation (A-9) holds, which means that $\widehat{u}_\tau$, the minimizer of $Q_{\tau,N}(u, \widehat{p})$, will be inside the ε interval around $\widetilde{u}_\tau$ with probability approaching 1, that is, $|\widehat{u}_\tau - \widetilde{u}_\tau| = o_p(1)$. *Q.E.D.*

Let us state the final result:

LEMMA A.5—Asymptotic Properties of $\widehat{q}_{1,\tau}$: *Under Assumptions* 1, 2, A.1, *and* A.2,

$$\sqrt{N}(\widehat{q}_{1,\tau} - q_{1,\tau}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \psi_{1,\tau}(Y_i, T_i, X_i) + o_p(1)$$

with $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \psi_{1,\tau}(Y_i, T_i, X_i) \xrightarrow{D} N(0, V_{1,\tau})$ *and where*

$$\begin{aligned} V_1 &= E[\psi_{1,\tau}^2(Y, T, X)] \\ &= E\left[\frac{V[g_{1,\tau}(Y)|X, T = 1]}{p(X)} + E^2[g_{1,\tau}(Y)|X, T = 1]\right]. \end{aligned}$$

PROOF: Defining $\widetilde{q}_{1,\tau} = \widetilde{u}_\tau/\sqrt{N} + q_{1,\tau}$, by Lemma A.4 we have

$$\begin{aligned} \sqrt{N}|\widehat{q}_{1,\tau} - \widetilde{q}_{1,\tau}| &= |\sqrt{N}(\widehat{q}_{1,\tau} - q_{1,\tau}) - \sqrt{N}(\widetilde{q}_{1,\tau} - q_{1,\tau})| \\ &\leq |\widehat{u}_\tau - \widetilde{u}_\tau| = o_p(1), \end{aligned}$$

that is, $\widehat{q}_{1,\tau}$ is asymptotically equivalent to $\widetilde{q}_{1,\tau}$ and Lemma A.5 follows immediately by Lemma A.3. *Q.E.D.*

The same result obtained for $q_{1,\tau}$ can be obtained analogously for $q_{0,\tau}$. In particular, with the same set of assumptions used in Lemma A.5, it is possible to derive an asymptotic linear influence function for $\widehat{q}_{0,\tau}$, $\psi_{0,\tau}$, which is analogous to $\psi_{1,\tau}$ and whose expression appears in Equation (7).

A consequence of Lemma A.5 is that $\widehat{\Delta}_\tau$, which is equal to the difference between $\widehat{q}_{1,\tau}$ and $\widehat{q}_{0,\tau}$, will have an asymptotically linear influence function and will be asymptotically normal. In other words, Theorem 1 will hold: $\sqrt{N}(\widehat{\Delta}_\tau - \Delta_\tau) = \frac{1}{\sqrt{N}}\sum_{i=1}^N \psi_\tau(Y_i, T_i, X_i) + o_p(1)$; with $\frac{1}{\sqrt{N}}\sum_{i=1}^N \psi_\tau(Y_i, T_i, X_i) \xrightarrow{D} N(0, V_\tau)$, where $\psi_\tau = \psi_{1,\tau} - \psi_{0,\tau}$ and

$$\begin{aligned} V_\tau &= E\big[(\psi_{1,\tau}(Y,T,X) - \psi_{0,\tau}(Y,T,X))^2\big] \\ &= E\left[\frac{V[g_{1,\tau}(Y)|X, T=1]}{p(X)} + \frac{V[g_{0,\tau}(Y)|X, T=0]}{1-p(X)}\right] \\ &\quad + E\big[\big(E[g_{1,\tau}(Y)|X, T=1] - E[g_{0,\tau}(Y)|X, T=0]\big)^2\big]. \qquad \textit{Q.E.D.} \end{aligned}$$

PROOF OF THEOREM 2: Here we state additional conditions and prove that $\widehat{V}_\tau$ is consistent for V_τ, the asymptotic variance of $\sqrt{N}\cdot\widehat{\Delta}_\tau$. Note that the same argument found in the following text could be extended easily to the estimation of $V_{\Delta_\tau|T=1}$, which is the asymptotic variance of $\sqrt{N}\cdot\widehat{\Delta}_{\tau|T=1}$, the quantile treatment effect on the treated.

ASSUMPTION A.3—Consistent Variance Estimation: *For $j = 0, 1$, the following conditions hold.*

Conditional Variance: (i) *probability* $\Pr(Y(j) \le q_{j,\tau}|X = x)$ *is bounded away from* 0 *and* 1, *and is* $s_{j,\tau}$*-times continuously differentiable, where* $s_{j,\tau} \ge 7r$ *and* r *is the dimension of* X; (ii) K_τ, *the order of* $H_{K_\tau}(x)$, *is of the form* $K_\tau = CN^{c_\tau}$, *where* C *is a constant and* $c_\tau \in (\frac{1}{4(s_{j,\tau}/r-1)}, \frac{1}{10})$.

Density:[10] (i) *The function* $f_j(\cdot)$ *is such that its second derivative*, $f_j''(\cdot)$, *is continuous*, *square integrable*, *and ultimately monotone*; (ii) h_j *is of the form* $h_j = CN^{-1/5}$, *where* C *is a constant and*, *therefore*, $\lim_{N\to+\infty} h_j = 0$ *and* $\lim_{N\to+\infty} N\cdot h_j = +\infty$; (iii) $\mathcal{K}_j(\cdot)$ *is a bounded probability density function that has finite fourth moment and symmetry about zero*.

We start by bounding

$$\begin{aligned} &|\widehat{V}_\tau - V_\tau| \\ &= \left|\frac{1}{N}\sum_{i=1}^N (\widehat{\varphi}_{\tau,i} + \widehat{\alpha}_{\tau,i})^2 - E\big[(\varphi_\tau(Y,T,X) + \alpha_\tau(Y,T,X))^2\big]\right| \end{aligned}$$

$$\le \left|\frac{1}{N}\sum_{i=1}^N (\widehat{\varphi}_{\tau,i} + \widehat{\alpha}_{\tau,i})^2 - (\widetilde{\varphi}_{\tau,i} + \widetilde{\alpha}_{\tau,i})^2\right| \tag{A-10}$$

$$+ \left|\frac{1}{N}\sum_{i=1}^N (\widetilde{\varphi}_{\tau,i} + \widetilde{\alpha}_{\tau,i})^2 - E\big[(\varphi_\tau(Y,T,X) + \alpha_\tau(Y,T,X))^2\big]\right|, \tag{A-11}$$

[10]These are standard textbook assumptions on kernel density estimation. See, for example, Wand and Jones (1995, Section 2.5).

where

$$\widetilde{\varphi}_{\tau,i} = \frac{T_i}{\widehat{p}(X_i)} \cdot g_{1,\tau}(Y_i) - \frac{1-T_i}{1-\widehat{p}(X_i)} \cdot g_{0,\tau}(Y_i),$$
$$\widetilde{\alpha}_{\tau,i} = -(T_i - \widehat{p}(X_i)) \cdot H_{K_\tau}(X_i)' \cdot \widetilde{\lambda}_\tau,$$

and

$$\widetilde{\lambda}_\tau = \left(\sum_{l=1}^{N} H_{K_\tau}(X_l) \cdot H_{K_\tau}(X_l)'\right)^{-1} \times \left(\sum_{i=1}^{N} H_{K_\tau}(X_i) \cdot \left(\frac{T_i \cdot g_{1,\tau}(Y_i)}{(\widehat{p}(X_i))^2} + \frac{(1-T_i) \cdot g_{0,\tau}(Y_i)}{(1-\widehat{p}(X_i))^2}\right)\right);$$

that is, unlike $\widehat{\varphi}_{\tau,i}$ and $\widehat{\alpha}_{\tau,i}$, the terms $\widetilde{\varphi}_{\tau,i}$ and $\widetilde{\alpha}_{\tau,i}$ use the true quantiles and the true density of potential outcomes. Subsequently, we show that such a bound converges to zero in probability.

Under Assumptions 1, 2, A.1, and A.2, expression (A-11) will be bounded by an $o_p(1)$ term. This is a consequence of a direct application of Theorem 2 of HIR, using $g_{1,\tau}(Y)$ and $g_{0,\tau}(Y)$, which are bounded functions of Y, instead of Y itself as in their case. We have then only to restrict attention to expression (A-10),

$$\left|\frac{1}{N}\sum_{i=1}^{N}(\widehat{\varphi}_{\tau,i} + \widehat{\alpha}_{\tau,i})^2 - (\widetilde{\varphi}_{\tau,i} + \widetilde{\alpha}_{\tau,i})^2\right| \leq V_A + V_B + V_C,$$

where $V_A = |\frac{1}{N}\sum_{i=1}^{N} \widehat{\varphi}^2_{\tau,i} - \widetilde{\varphi}^2_{\tau,i}|$, $V_B = |\frac{1}{N}\sum_{i=1}^{N} \widehat{\alpha}^2_{\tau,i} - \widetilde{\alpha}^2_{\tau,i}|$, and $V_C = |\frac{2}{N}\sum_{i=1}^{N} \widehat{\varphi}_{\tau,i} \cdot \widehat{\alpha}_{\tau,i} - \widetilde{\varphi}_{\tau,i} \cdot \widetilde{\alpha}_{\tau,i}|$.

Now, term-by-term,

$$V_A \leq \left|\frac{1}{N}\sum_{i=1}^{N} \frac{p(X_i)}{\widehat{p}^2(X_i)} \cdot \frac{T_i}{p(X_i)} \cdot (\widehat{g}^2_{1,\tau}(Y_i) - g^2_{1,\tau}(Y_i))\right| + \left|\frac{1}{N}\sum_{i=1}^{N} \frac{1-p(X_i)}{(1-\widehat{p}(X_i))^2} \cdot \frac{1-T_i}{1-p(X_i)} \cdot (\widehat{g}^2_{0,\tau}(Y_i) - g^2_{0,\tau}(Y_i))\right|$$

because the cross-product term will be zero because it will involve $T_i \cdot (1-T_i) = 0$ for all i. The first term of the foregoing sum will be

$$\text{(A-12)} \qquad \left|\frac{1}{N}\sum_{i=1}^{N} \frac{p(X_i)}{\widehat{p}^2(X_i)} \cdot \frac{T_i}{p(X_i)} \cdot (\widehat{g}^2_{1,\tau}(Y_i) - g^2_{1,\tau}(Y_i))\right|$$

$$\leq \frac{\sup_{x\in\mathcal{X}}(p(x))}{\inf_{x\in\mathcal{X}}(\widehat{p}^2(x))}\cdot\frac{1}{N}\sum_{i=1}^{N}\left|\frac{T_i}{p(X_i)}\cdot(\widehat{g}^2_{1,\tau}(Y_i)-g^2_{1,\tau}(Y_i))\right|$$
$$=\frac{C}{N}\sum_{i=1}^{N}\left|\frac{T_i}{p(X_i)}\cdot(\widehat{g}^2_{1,\tau}(Y_i)-g^2_{1,\tau}(Y_i))\right|,$$

where C is a generic constant. By definition, expression (A-12) is equal to

$$\frac{C}{N}\sum_{i=1}^{N}\left|\frac{T_i}{p(X_i)}\cdot\left(\frac{(\mathbb{1}\{Y_i\leq q_{1,\tau}\}-\tau)^2}{f_1^2(q_{1,\tau})}-\frac{(\mathbb{1}\{Y_i\leq \widehat{q}_{1,\tau}\}-\tau)^2}{\widehat{f}_1^2(\widehat{q}_{1,\tau})}\right)\right|$$
$$\leq (f_1(q_{1,\tau})\cdot\widehat{f}_1(\widehat{q}_{1,\tau}))^{-2}\cdot|\widehat{f}_1^2(\widehat{q}_{1,\tau})-f_1^2(q_{1,\tau})|$$
$$\times\frac{C}{N}\sum_{i=1}^{N}\left|\frac{T_i}{p(X_i)}\cdot\mathbb{1}\{Y_i\leq q_{1,\tau}\}\right|$$
$$+(f_1(q_{1,\tau})\cdot\widehat{f}_1(\widehat{q}_{1,\tau}))^{-2}\cdot f_1^2(q_{1,\tau})$$
$$\times\frac{C}{N}\sum_{i=1}^{N}\left|\frac{T_i}{p(X_i)}\cdot(\mathbb{1}\{Y_i\leq q_{1,\tau}\}-\mathbb{1}\{Y_i\leq \widehat{q}_{1,\tau}\})\right|,$$

where C is a generic constant. Now we show that $\widehat{f}_1(y)-f_1(y)=o_p(1)$ as the sample size $N\to\infty$ and the bandwidth for density estimation $h_1\to 0$. Define the auxiliary quantity based on true propensity score:

$$\widetilde{f}_1(y)=\frac{1}{N}\sum_{i=1}^{N}\frac{T_i}{p(X_i)}\frac{\mathcal{K}_1\left(\frac{Y_i-y}{h_1}\right)}{h_1}.$$

Then

$$|\widehat{f}_1(y)-f_1(y)|$$
$$=\left|\frac{1}{N}\sum_{i=1}^{N}\frac{T_i}{\widehat{p}(X_i)}\frac{\mathcal{K}_1\left(\frac{Y_i-y}{h_1}\right)}{h_1}-f_1(y)\right|$$
$$\leq\left|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{T_i}{\widehat{p}(X_i)}-\frac{T_i}{p(X_i)}\right)\cdot\frac{\mathcal{K}_1\left(\frac{Y_i-y}{h_1}\right)}{h_1}\right|$$
$$+\left|\widetilde{f}_1(y)-E[\widetilde{f}_1(y)]\right|+\left|E[\widetilde{f}_1(y)]-f_1(y)\right|.$$

The last two terms of the sum are typically encountered in kernel density estimation, the key difference being that $\widetilde{f}_1(y)$ is a weighted kernel density estimator. Given the standard assumptions embedded in Assumption A.3(II), and the ignorability and common support assumptions,[11] we can invoke known results for the order of the last two terms of the preceding sum (as in, for example, Wand and Jones (1995)): $|\widetilde{f}_1(y) - E[\widetilde{f}_1(y)]| = O_p((Nh_1)^{-1})$ and $|E[\widetilde{f}_1(y)] - f_1(y)| = O(h_1^2)$.

Now consider

$$\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{T_i}{\widehat{p}(X_i)} - \frac{T_i}{p(X_i)}\right)\cdot\frac{\mathcal{K}_1\left(\frac{Y_i-y}{h_1}\right)}{h_1}\right| \\
&\quad\leq \frac{\sup_{x\in\mathcal{X}}(|p(x)-\widehat{p}(x)|)}{\inf_{x\in\mathcal{X}}(\widehat{p}(x))}\cdot\left|\frac{1}{N}\sum_{i=1}^{N}\frac{T_i}{p(X_i)}\cdot\frac{\mathcal{K}_1\left(\frac{Y_i-y}{h_1}\right)}{h_1}\right| \\
&\quad\leq \left(C_1\cdot N^{(1-s/2r)c} + C_2\cdot N^{(3c-1)/2}\cdot O_p(1)\right)\cdot\left(O_p((Nh_1)^{-1}) + O(h_1^2)\right) \\
&\quad= o_p(1)\cdot o_p(1) = o_p(1).
\end{aligned}$$

It is thus clear that the weighted kernel density estimator $\widehat{f}_1(y)$, which uses nonparametric weights (given by $T_i\cdot(N\cdot\widehat{p}(X_i))^{-1}$), will be consistent for $f_1(y)$. More formally, for a bandwidth sequence $h_1(N)$ such that as $N\to\infty$, $h_1(N)\to 0$ and $N\cdot h_1(N)\to\infty$, $|\widehat{f}_1(y) - f_1(y)| = O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) = o_p(1)$. That result guarantees that

$$\begin{aligned}
&\frac{C}{N}\sum_{i=1}^{N}\left|\frac{T_i}{p(X_i)}\cdot(\widehat{g}^2_{1,\tau}(Y_i) - g^2_{1,\tau}(Y_i))\right| \\
&\quad\leq (f_1(q_{1,\tau})\cdot\widehat{f}_1(\widehat{q}_{1,\tau}))^{-2}\cdot|\widehat{f}_1^2(\widehat{q}_{1,\tau}) - f_1^2(q_{1,\tau})| \\
&\qquad\times\frac{C_1}{N}\sum_{i=1}^{N}\left|\frac{T_i}{p(X_i)}\cdot\mathbb{1}\{Y_i\leq q_{1,\tau}\}\right| \\
&\qquad+ (f_1(q_{1,\tau})\cdot\widehat{f}_1(\widehat{q}_{1,\tau}))^{-2}\cdot f_1^2(q_{1,\tau}) \\
&\qquad\times\frac{C_2}{N}\sum_{i=1}^{N}\left|\frac{T_i}{p(X_i)}\cdot\left(\mathbb{1}\{Y_i\leq q_{1,\tau}\} - \mathbb{1}\{Y_i\leq\widehat{q}_{1,\tau}\}\right)\right| \\
&\quad\leq C_1\cdot|\widehat{f}_1(\widehat{q}_{1,\tau}) - f_1(q_{1,\tau})|\cdot|\widehat{f}_1(\widehat{q}_{1,\tau}) + f_1(q_{1,\tau})| \\
&\qquad\times O_p\left(\sqrt{E\left[\left(\mathbb{1}\{Y(1)\leq q_{1,\tau}\}\right)^2\right]}\right)
\end{aligned}$$

[11]The ignorability and common support assumptions guarantee that $E[\frac{T}{p(X)}\cdot\mathcal{K}_1(\frac{Y-y}{h_1})] = E[\mathcal{K}_1(\frac{Y(1)-y}{h_1})]$ for a fixed h_1.

$$
\begin{aligned}
&\quad + C_2 \cdot |\widehat{q}_{1,\tau} - q_{1,\tau}| \cdot O_p\left(\sqrt{E\left[\left(\frac{T}{p(X)}\right)^2\right]}\right) \\
&\le C_1 \cdot \big(|\widehat{f}_1(\widehat{q}_{1,\tau}) - f_1(\widehat{q}_{1,\tau})| + |f_1(\widehat{q}_{1,\tau}) - f_1(q_{1,\tau})|\big) \\
&\quad \times \big(|\widehat{f}_1(\widehat{q}_{1,\tau}) - f_1(\widehat{q}_{1,\tau})| + |f_1(\widehat{q}_{1,\tau}) + f_1(q_{1,\tau})|\big) \cdot O_p(1) \\
&\quad + C_2 \cdot O_p(N^{-1/2}) \cdot O_p(1) \\
&= \big(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2})\big) \\
&\quad \times (o_p(1) + O_p(1)) \cdot O_p(1) + O_p(N^{-1/2}) \\
&= O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2}) = o_p(1).
\end{aligned}
$$

A simple analogy allows us to conclude that

$$
\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^{N} \frac{1 - p(X_i)}{(1 - \widehat{p}(X_i))^2} \cdot \frac{1 - T_i}{1 - p(X_i)} \cdot (\widehat{g}_{0,\tau}^2(Y_i) - g_{0,\tau}^2(Y_i))\right| \\
&\quad = O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2}) = o_p(1),
\end{aligned}
$$

which allows us to write

$$
V_A = O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2}) = o_p(1).
$$

The second term of the original sum, V_B, will also be bounded. To see this, note that

$$
\begin{aligned}
V_B &= \left|\frac{1}{N}\sum_{i=1}^{N}(T_i - \widehat{p}(X_i))^2\right. \\
&\qquad \left. \times \big(\widehat{\lambda}_\tau' \cdot H_{K_\tau}(X_i) \cdot H_{K_\tau}(X_i)' \cdot \widehat{\lambda}_\tau - \widetilde{\lambda}_\tau' \cdot H_{K_\tau}(X_i) \cdot H_{K_\tau}(X_i)' \cdot \widetilde{\lambda}_\tau\big)\right| \\
&\le \left|(\widehat{\lambda}_\tau - \widetilde{\lambda}_\tau)' \cdot \frac{1}{N}\sum_{i=1}^{N}(T_i - \widehat{p}(X_i))^2 \cdot H_{K_\tau}(X_i) \cdot H_{K_\tau}(X_i)' \cdot \widehat{\lambda}_\tau\right| \\
&\quad + \left|\widetilde{\lambda}_\tau' \cdot \frac{1}{N}\sum_{i=1}^{N}(T_i - \widehat{p}(X_i))^2 \cdot H_{K_\tau}(X_i) \cdot H_{K_\tau}(X_i)' \cdot (\widehat{\lambda}_\tau - \widetilde{\lambda}_\tau)\right| \\
&\le \|(\widehat{\lambda}_\tau - \widetilde{\lambda}_\tau)\| \cdot (\|\widehat{\lambda}_\tau\| + \|\widetilde{\lambda}_\tau\|) \cdot \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x) \cdot H_{K_\tau}(x)'\| \\
&\quad \times \left|\frac{1}{N}\sum_{i=1}^{N}(T_i - p(X_i) + p(X_i) - \widehat{p}(X_i))^2\right|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \|(\widehat{\lambda}_\tau - \widetilde{\lambda}_\tau)\| \cdot (\|\widehat{\lambda}_\tau\| + \|\widetilde{\lambda}_\tau\|) \cdot \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\|^2 \\
&\quad \times \Bigg(\left| \frac{1}{N} \sum_{i=1}^{N} (T_i - p(X_i))^2 \right| + \sup_{x\in\mathcal{X}} (p(x) - \widehat{p}(x))^2 \\
&\qquad + C \cdot \sup_{x\in\mathcal{X}} \big(|p(x) - \widehat{p}(x)|\big) \cdot \left| \frac{1}{N} \sum_{i=1}^{N} T_i - p(X_i) \right| \Bigg) \\
&\leq C_1 \cdot \|(\widehat{\lambda}_\tau - \widetilde{\lambda}_\tau)\| \cdot (\|\widehat{\lambda}_\tau\| + \|\widetilde{\lambda}_\tau\|) \cdot \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\|^2 \\
&\quad \times \Big(O_p\big(\sqrt{E\big[(T - p(X))^4\big]}\big) \\
&\qquad + o_p(1) \cdot \big(O_p(1) + O_p\big(\sqrt{E\big[(T - p(X))^2\big]}\big)\big)\Big) \\
&\leq C \cdot \|(\widehat{\lambda}_\tau - \widetilde{\lambda}_\tau)\| \cdot (\|\widehat{\lambda}_\tau\| + \|\widetilde{\lambda}_\tau\|) \cdot \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\|^2 \cdot O_p(1).
\end{aligned}
$$

Now, let us concentrate on

$$
\begin{aligned}
&\|(\widehat{\lambda}_\tau - \widetilde{\lambda}_\tau)\| \\
&\quad \leq \inf_{x\in\mathcal{X}} \|H_{K_\tau}(x)\|^{-2} \\
&\qquad \times \Bigg\| \Bigg(\frac{1}{N} \sum_{i=1}^{N} H_{K_\tau}(X_i) \cdot \bigg(\frac{T_i \cdot (\widehat{g}_{1,\tau}(Y_i) - g_{1,\tau}(Y_i))}{(\widehat{p}(X_i))^2} \\
&\qquad\qquad + \frac{(1 - T_i) \cdot (\widehat{g}_{0,\tau}(Y_i) - g_{0,\tau}(Y_i))}{(1 - \widehat{p}(X_i))^2} \bigg) \Bigg) \Bigg\| \\
&\quad \leq C \cdot \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\| \cdot \Bigg(\left| \frac{1}{N} \sum_{i=1}^{N} \frac{T_i \cdot (\widehat{g}_{1,\tau}(Y_i) - g_{1,\tau}(Y_i))}{(\widehat{p}(X_i))^2} \right| \\
&\qquad\qquad + \left| \frac{1}{N} \sum_{i=1}^{N} \frac{(1 - T_i) \cdot (\widehat{g}_{0,\tau}(Y_i) - g_{0,\tau}(Y_i))}{(1 - \widehat{p}(X_i))^2} \right| \Bigg) \\
&\quad \leq C_1 \cdot \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\| \cdot \frac{\sup_{x\in\mathcal{X}}(p(x))}{\inf_{x\in\mathcal{X}}(\widehat{p}^2(x))} \\
&\qquad \times \left| \frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{p(X_i)} \cdot (\widehat{g}_{1,\tau}(Y_i) - g_{1,\tau}(Y_i)) \right| \\
&\qquad + C_2 \cdot \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\| \cdot \frac{1 - \inf_{x\in\mathcal{X}}(p(x))}{1 - \sup_{x\in\mathcal{X}}(\widehat{p}^2(x))}
\end{aligned}
$$

$$\times \left| \frac{1}{N} \sum_{i=1}^{N} \frac{(1-T_i)}{1-p(X_i)} \cdot (\widehat{g}_{0,\tau}(Y_i) - g_{0,\tau}(Y_i)) \right|$$

and, following the arguments to bound expression (A-12), we have that

$$\begin{aligned}
&\left| \frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{p(X_i)} \cdot (\widehat{g}_{1,\tau}(Y_i) - g_{1,\tau}(Y_i)) \right| \\
&\quad \leq \frac{1}{N} \sum_{i=1}^{N} \left| \frac{T_i}{p(X_i)} \cdot \left(\frac{\mathbb{1}\{Y_i \leq q_{1,\tau}\} - \tau}{f_1(q_{1,\tau})} - \frac{\mathbb{1}\{Y_i \leq \widehat{q}_{1,\tau}\} - \tau}{\widehat{f}_1(\widehat{q}_{1,\tau})} \right) \right| \\
&\quad \leq C \cdot (f_1(q_{1,\tau}) \cdot \widehat{f}_1(\widehat{q}_{1,\tau}))^{-1} \cdot |\widehat{f}_1(\widehat{q}_{1,\tau}) - f_1(q_{1,\tau})| \\
&\qquad \times \left(\frac{1}{N} \sum_{i=1}^{N} \left| \frac{T_i}{p(X_i)} \cdot (\mathbb{1}\{Y_i \leq q_{1,\tau}\} - \tau) \right| \right. \\
&\qquad\qquad \left. + f_1(q_{1,\tau}) \cdot \frac{1}{N} \sum_{i=1}^{N} |\mathbb{1}\{Y_i \leq \widehat{q}_{1,\tau}\} - \mathbb{1}\{Y_i \leq q_{1,\tau}\}| \right) \\
&\quad \leq C \cdot \left(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2}) \right) \\
&\qquad \times (O_p(1) + O_p(N^{-1/2})) \\
&\quad = O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2}) = o_p(1).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&C \cdot \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x)\| \cdot \frac{\sup_{x \in \mathcal{X}}(p(x))}{\inf_{x \in \mathcal{X}}(\widehat{p}^2(x))} \\
&\qquad \times \left| \frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{p(X_i)} \cdot (\widehat{g}_{1,\tau}(Y_i) - g_{1,\tau}(Y_i)) \right| \\
&\quad = C \cdot \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x)\| \cdot \left(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2}) \right) \\
&\quad = \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x)\| \cdot o_p(1)
\end{aligned}$$

and by analogy,

$$\begin{aligned}
&C \cdot \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x)\| \cdot \frac{1 - \inf_{x \in \mathcal{X}}(p(x))}{1 - \sup_{x \in \mathcal{X}}(\widehat{p}^2(x))} \\
&\qquad \times \left| \frac{1}{N} \sum_{i=1}^{N} \frac{(1-T_i)}{1-p(X_i)} \cdot (\widehat{g}_{0,\tau}(Y_i) - g_{0,\tau}(Y_i)) \right|
\end{aligned}$$

$$= C \cdot \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x)\| \cdot \big(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2})\big)$$
$$= \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x)\| \cdot o_p(1),$$

so

$$\begin{aligned}
&\|(\widehat{\lambda}_\tau - \widetilde{\lambda}_\tau)\| \\
&\quad = C \cdot \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x)\| \cdot \big(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2})\big) \\
&\quad = \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x)\| \cdot o_p(1)
\end{aligned}$$

and

$$\begin{aligned}
&C \cdot \|(\widehat{\lambda}_\tau - \widetilde{\lambda}_\tau)\| \cdot (\|\widehat{\lambda}_\tau\| + \|\widetilde{\lambda}_\tau\|) \cdot \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x)\|^2 \cdot O_p(1) \\
&\quad = C \cdot (\|\widehat{\lambda}_\tau\| + \|\widetilde{\lambda}_\tau\|) \cdot \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x)\|^3 \\
&\qquad \times \big(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2})\big) \\
&\quad \le C \cdot \big(\|(\widehat{\lambda}_\tau - \widetilde{\lambda}_\tau)\| + 2\|\widetilde{\lambda}_\tau\|\big) \cdot \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x)\|^3 \\
&\qquad \times \big(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2})\big) \\
&\quad \le C_1 \cdot \|\widetilde{\lambda}_\tau\| \cdot \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x)\|^3 \\
&\qquad \times \big(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2})\big) \\
&\qquad + C_2 \cdot \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x)\|^4 \\
&\qquad\quad \times \big(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2})\big)^2.
\end{aligned}$$

However,

$$\begin{aligned}
\|\widetilde{\lambda}_\tau\| &= \Bigg\| \left(\frac{1}{N} \sum_{l=1}^N H_{K_\tau}(X_l) \cdot H_{K_\tau}(X_l)' \right)^{-1} \\
&\qquad \times \left(\frac{1}{N} \sum_{i=1}^N H_{K_\tau}(X_i) \cdot \left(\frac{T_i \cdot g_{1,\tau}(Y_i)}{(\widehat{p}(X_i))^2} + \frac{(1 - T_i) \cdot g_{0,\tau}(Y_i)}{(1 - \widehat{p}(X_i))^2} \right) \right) \Bigg\| \\
&\quad \le C_1 \cdot \sup_{x \in \mathcal{X}} \|H_{K_\tau}(x)\| \cdot \frac{\sup_{x \in \mathcal{X}} (p(x))}{\inf_{x \in \mathcal{X}} (\widehat{p}^2(x))} \cdot \left| \frac{1}{N} \sum_{i=1}^N \frac{T_i}{p(X_i)} \cdot g_{1,\tau}(Y_i) \right|
\end{aligned}$$

$$
\begin{aligned}
&+ C_2 \cdot \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\| \cdot \frac{1-\inf_{x\in\mathcal{X}}(p(x))}{1-\sup_{x\in\mathcal{X}}(\widehat{p}^2(x))} \\
&\quad \times \left| \frac{1}{N}\sum_{i=1}^{N} \frac{(1-T_i)}{1-p(X_i)} \cdot g_{0,\tau}(Y_i) \right| \\
&\le \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\| \\
&\quad \times \left(C_1 \cdot O_p\left(\sqrt{E\left[g_{1,\tau}^2(Y(1))\right]}\right) + C_2 \cdot O_p\left(\sqrt{E\left[g_{0,\tau}^2(Y(0))\right]}\right)\right) \\
&\le C \cdot \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\| \cdot O_p(1)
\end{aligned}
$$

and then

$$
\begin{aligned}
V_B &\le C_1 \cdot \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\|^4 \\
&\quad \times \left(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2})\right) \\
&\quad + C_2 \cdot \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\|^4 \\
&\quad\quad \times \left(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2})\right)^2 \\
&\le C \cdot \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\|^4 \\
&\quad \times \left(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2})\right) \\
&\le C \cdot O(N^{4c_\tau}) \cdot \left(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2})\right) \\
&= O(N^{4c_\tau}) \cdot \left(O_p(N^{-4/5}) + O(N^{-2/5}) + o_p(1) + O_p(N^{-1/2})\right) \\
&= O_p(N^{4c_\tau - 2/5}) = o_p(1)
\end{aligned}
$$

because $c_\tau \in (\frac{1}{4(s_{j,\tau}/r-1)}, \frac{1}{10})$ and $s_{j,\tau} \ge 7r$.

The last term to bound is

$$
\begin{aligned}
V_C &= \left| \frac{2}{N}\sum_{i=1}^{N} \widehat{\varphi}_{\tau,i} \cdot \widehat{\alpha}_{\tau,i} - \widetilde{\varphi}_{\tau,i} \cdot \widetilde{\alpha}_{\tau,i} \right| \\
&\le C_1 \left| \frac{1}{N}\sum_{i=1}^{N} \widetilde{\varphi}_{\tau,i} \cdot (\widehat{\alpha}_{\tau,i} - \widetilde{\alpha}_{\tau,i}) \right| + C_2 \left| \frac{1}{N}\sum_{i=1}^{N} (\widehat{\varphi}_{\tau,i} - \widetilde{\varphi}_{\tau,i}) \cdot \widehat{\alpha}_{\tau,i} \right|.
\end{aligned}
$$

Focusing on the first part of that sum,

$$
\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^{N}\widetilde{\varphi}_{\tau,i}\cdot(\widehat{\alpha}_{\tau,i}-\widetilde{\alpha}_{\tau,i})\right| \\
&\quad=\left|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{T_i}{\widehat{p}(X_i)}\cdot g_{1,\tau}(Y_i)-\frac{1-T_i}{1-\widehat{p}(X_i)}\cdot g_{0,\tau}(Y_i)\right)\right. \\
&\qquad\quad\left.\times(T_i-\widehat{p}(X_i))\cdot H_{K_\tau}(X_i)'\cdot(\widetilde{\lambda}_\tau-\widehat{\lambda}_\tau)\right| \\
&\quad\leq\sup_{x\in\mathcal{X}}\|H_{K_\tau}(x)\|\cdot\|\widehat{\lambda}_\tau-\widetilde{\lambda}_\tau\|\cdot\frac{1-\inf_{x\in\mathcal{X}}(p(x))\cdot\sup_{x\in\mathcal{X}}(p(x))}{\inf_{x\in\mathcal{X}}(\widehat{p}(x))} \\
&\qquad\times\left|\frac{1}{N}\sum_{i=1}^{N}\frac{T_i}{p(X_i)}\cdot g_{1,\tau}(Y_i)\right| \\
&\qquad+\sup_{x\in\mathcal{X}}\|H_{K_\tau}(x)\|\cdot\|\widehat{\lambda}_\tau-\widetilde{\lambda}_\tau\|\cdot\frac{1-\inf_{x\in\mathcal{X}}(p(x))\cdot\sup_{x\in\mathcal{X}}(\widehat{p}(x))}{1-\sup_{x\in\mathcal{X}}(\widehat{p}(x))} \\
&\qquad\quad\times\left|\frac{1}{N}\sum_{i=1}^{N}\frac{1-T_i}{1-p(X_i)}\cdot g_{0,\tau}(Y_i)\right| \\
&\quad\leq C_1\cdot\sup_{x\in\mathcal{X}}\|H_{K_\tau}(x)\|^2 \\
&\qquad\times\big(O_p((Nh_1)^{-1})+O(h_1^2)+o_p(1)+O_p(N^{-1/2})\big)\cdot O_p(1) \\
&\qquad+C_2\cdot\sup_{x\in\mathcal{X}}\|H_{K_\tau}(x)\|^2 \\
&\qquad\quad\times\big(O_p((Nh_1)^{-1})+O(h_1^2)+o_p(1)+O_p(N^{-1/2})\big)\cdot O_p(1) \\
&\quad\leq C\cdot O(N^{2c_\tau})\cdot\big(O_p((Nh_1)^{-1})+O(h_1^2)+o_p(1)+O_p(N^{-1/2})\big) \\
&\quad=O(N^{2c_\tau})\cdot\big(O_p(N^{-4/5})+O(N^{-2/5})+o_p(1)+O_p(N^{-1/2})\big) \\
&\quad=O_p(N^{2c_\tau-2/5})=o_p(1).
\end{aligned}
$$

The second part is,

$$
\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^{N}(\widehat{\varphi}_{\tau,i}-\widetilde{\varphi}_{\tau,i})\cdot\widehat{\alpha}_{\tau,i}\right| \\
&\quad=\left|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{T_i}{\widehat{p}(X_i)}\cdot(\widehat{g}_{1,\tau}(Y_i)-g_{1,\tau}(Y_i))\right.\right.
\end{aligned}
$$

$$
\begin{aligned}
&\qquad\qquad - \frac{1-T_i}{1-\widehat{p}(X_i)} \cdot (\widehat{g}_{0,\tau}(Y_i) - g_{0,\tau}(Y_i))\Big) \\
&\quad \times (T_i - \widehat{p}(X_i)) \cdot H_{K_\tau}(X_i)' \cdot \widehat{\lambda}_\tau \Bigg| \\
&\le C_1 \cdot \|\widehat{\lambda}_\tau\| \cdot \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\| \cdot \left| \frac{1}{N} \sum_{i=1}^{N} \frac{T_i}{p(X_i)} \cdot (\widehat{g}_{1,\tau}(Y_i) - g_{1,\tau}(Y_i)) \right| \\
&\quad + C_2 \cdot \|\widehat{\lambda}_\tau\| \cdot \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\| \\
&\qquad \times \left| \frac{1}{N} \sum_{i=1}^{N} \frac{1-T_i}{1-p(X_i)} \cdot (\widehat{g}_{0,\tau}(Y_i) - g_{0,\tau}(Y_i)) \right| \\
&\le C \cdot \big(\|(\widehat{\lambda}_\tau - \widetilde{\lambda}_\tau)\| + \|\widetilde{\lambda}\| \big) \\
&\quad \times \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\| \cdot \big(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2}) \big) \\
&\le C \cdot (o_p(1) + O_p(1)) \cdot \sup_{x\in\mathcal{X}} \|H_{K_\tau}(x)\|^2 \\
&\quad \times \big(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2}) \big) \\
&\le O(N^{2c_\tau}) \cdot \big(O_p((Nh_1)^{-1}) + O(h_1^2) + o_p(1) + O_p(N^{-1/2}) \big) = o_p(1),
\end{aligned}
$$

so, $V_C = o_p(1)$. This concludes the proof because

$$
|\widehat{V}_\tau - V_\tau| \le V_A + V_B + V_C + o_p(1) = o_p(1). \qquad \textit{Q.E.D.}
$$

PROOF OF THEOREM 3: This proof is an extension to the quantile case of the proofs by Hahn (1998) and by HIR for the mean case. Both references use the machinery presented by Bickel, Klaassen, Ritov, and Wellner (1993) and Newey (1990, 1994). We start by defining the densities, with respect to some σ-finite measure, of $(Y(1), Y(0), T, X)$ and of the observed data (Y, T, X). Under Assumption 1 both densities represent the same statistical model and are, therefore, equivalent. These densities can be written as $f(y(1), y(0), t, x) = f(y(1), y(0)|x)p(x)^t(1-p(x))^{1-t}f(x)$ and $f(y, t, x) = [f_1(y|x)p(x)]^t[f_0(y|x)(1-p(x))]^{1-t}f(x)$, where $f_1(y|x) = \int f(y, z|x)\,dz$ and $f_0(y|x) = \int f(z, y|x)\,dz$. Working with the density of observed data, consider the regular parametric submodel indexed by θ, a finite dimensional vector:

$$
f(y, t, x|\theta) = [f_1(y|x;\theta)p(x|\theta)]^t \big[f_0(y|x;\theta)(1-p(x|\theta))\big]^{1-t} f(x|\theta).
$$

By a normalization argument, let $f(y,t,x) = f(y,t,x|\theta_0)$. The score of a parametric submodel indexed by θ is given by

$$s(y,t,x|\theta) = ts_1(y|x;\theta) + (1-t)s_0(y|x;\theta) + \frac{t - p(x|\theta)}{p(x|\theta)(1-p(x|\theta))} p'(x|\theta) + s_x(x|\theta),$$

where, for $j = 0, 1$, $s_j(y|x;\theta) = \frac{\partial}{\partial\theta}\log f_j(y|x;\theta)$; $p'(x|\theta) = \frac{\partial}{\partial\theta} p(x|\theta)$, and $s_x(x|\theta) = \frac{\partial}{\partial\theta}\log f(x|\theta)$. Again we normalize: $s(y,t,x) = s(y,t,x|\theta_0)$. To find the efficient influence functions of the parameters of interest, $\Delta_\tau(\theta)$ and $\Delta_{\tau|T=1}(\theta)$, we need first to define the tangent space of this statistical model. This will be the set $\mathcal{S}$ of all possible score functions and it is defined as

$$\begin{aligned}\mathcal{S} = \big\{ & S : \mathbb{R} \times \{0,1\} \times \mathcal{X} \to \mathbb{R} | \\ & S(y,t,x) = ts_1(y|x) + (1-t)s_0(y|x) + a(x)(t - p(x)) + s_x(x); \\ & \text{and } E[s_j(Y|X)|X = x, T = j] = E[s_x(X)] = 0, \\ & \forall\, x \text{ and } j = 0, 1\big\},\end{aligned}$$

where $a(x)$ is some square-integrable measurable function of x. Next we show that both $\Delta_\tau(\theta)$ and $\Delta_{\tau|T=1}(\theta)$ are pathwise differentiable, that is, we show that for each one the derivative with respect to θ evaluated at θ_0 is equal to the expectation of the product of the score $s(Y,T,X)$ and the respective efficient influence functions $\psi_\tau(Y,T,X)$ and $\psi_{\tau|T=1}(Y,T,X)$, to be defined. After we show pathwise differentiability, we find the projection of the influence function on the set of scores. That projection is often called the efficient influence function. If an influence function belongs to the set $\mathcal{S}$, then its projection onto $\mathcal{S}$ is the original influence function itself. Therefore, the goal is to find an influence function that already belongs to the set of scores. A function that is in the set of scores must be written as

$$\psi = Tc_1(Y,X) + (1-T)c_0(Y,X) + a(X)(T - p(X)) + c_x(X),$$

where $E[c_j(Y,X)|X = x, T = j] = E[c_x(X)] = 0\ \forall\, x$ and $j = 0, 1$. Starting with $q_{1,\tau}$, the first part of the parameter is Δ_τ. For the parametric submodel indexed by θ, we have, for all θ, $0 = \iint (\mathbb{1}\{y \le q_{1,\tau}(\theta)\} - \tau) f_1(y|x;\theta) f(x|\theta)\, dy\, dx$. Thus, using the normalization $q_{1,\tau} = q_{1,\tau}(\theta_0)$ and by an application of Leibnitz's rule, we have

$$\frac{\partial q_{1,\tau}(\theta_0)}{\partial\theta} = \iint g_{1,\tau}(y) s_1(y|x) f_1(y|x) f(x)\, dy\, dx + \iint g_{1,\tau}(y) s_x(x) f_1(y|x) f(x)\, dy\, dx,$$

where the function $g_{j,\tau}$, $j = 0, 1$, was defined in Equation (8). After similar calculations for $q_{0,\tau}$, we can express the derivative of $\Delta_\tau(\theta)$ evaluated at θ_0 as

$$\begin{aligned}\text{(A-13)}\quad \frac{\partial \Delta_\tau(\theta_0)}{\partial\theta} &= \iint g_{1,\tau}(y)s_1(y|x)f_1(y|x)f(x)\,dy\,dx \\ &\quad - \iint g_{0,\tau}(y)s_0(y|x)f_0(y|x)f(x)\,dy\,dx \\ &\quad + \iint g_{1,\tau}(y)s_x(x)f_1(y|x)f(x)\,dy\,dx \\ &\quad - \iint g_{0,\tau}(y)s_x(x)f_0(y|x)f(x)\,dy\,dx.\end{aligned}$$

The next goal is to find a function of (Y, T, X) such that the expectation of the product of that function times the score is equal to Equation (A-13). A solution to this problem is

$$\begin{aligned}\psi_\tau(Y, T, X) &= \frac{T(g_{1,\tau}(Y) - E[g_{1,\tau}(Y)|X, T = 1])}{p(X)} \\ &\quad - \frac{(1 - T)(g_{0,\tau}(Y) - E[g_{0,\tau}(Y)|X, T = 0])}{1 - p(X)} \\ &\quad + E[g_{1,\tau}(Y)|X, T = 1] - E[g_{0,\tau}(Y)|X, T = 0].\end{aligned}$$

Note, however, that by inspection this influence function belongs to the set of scores. Hence, ψ_τ is the efficient influence function and because of that its variance is equal to $E[\psi_\tau^2(Y, T, X)]$, which is the semiparametric efficiency bound for Δ_τ, V_τ.

Now we do the same for $\Delta_{\tau|T=1}$. For a parametric submodel indexed by θ, we have

$$\begin{aligned}0 = \iint &\frac{p(x|\theta)}{\int p(x|\theta)f(x|\theta)\,dx} \\ &\times \big(\mathbb{1}\{y \le q_{1,\tau|T=1}(\theta)\} - \tau\big) f_1(y|x;\theta)f(x|\theta)\,dy\,dx.\end{aligned}$$

Again we normalize: $q_{1,\tau|T=1} = q_{1,\tau|T=1}(\theta_0)$. The derivative evaluated at θ_0 is equal to

$$\begin{aligned}\frac{\partial q_{1,\tau|T=1}(\theta_0)}{\partial\theta} &= \iint g_{1,\tau_{|T=1}}(y)p(x)s_1(y|x)f_1(y|x)f(x)\,dy\,dx \\ &\quad + \int E[g_{1,\tau_{|T=1}}(Y)|X = x]p'(x)f(x)\,dx \\ &\quad + \int E[g_{1,\tau_{|T=1}}(Y)|X = x]p(x)s_x(x)f(x)\,dx,\end{aligned}$$

where the function $g_{1,\tau|T=1}(j)$, $j=0,1$, was defined by Equation (10). Whereas the same sorts of calculations are true for $q_{0,\tau|T=1}$, we can express the derivative of $\Delta_{\tau|T=1}(\theta)$ evaluated at θ_0 as

$$\begin{aligned}\frac{\partial \Delta_{\tau|T=1}(\theta_0)}{\partial\theta} &= \iint g_{1,\tau|T=1}(y)\,p(x)s_1(y|x)f_1(y|x)f(x)\,dy\,dx\\ &\quad - \iint g_{0,\tau|T=1}(y)\,p(x)s_0(y|x)f_0(y|x)f(x)\,dy\,dx\\ &\quad + \int E[g_{1,\tau|T=1}(Y)|X=x]p'(x)f(x)\,dx\\ &\quad - \int E[g_{0,\tau|T=1}(Y)|X=x]p'(x)f(x)\,dx\\ &\quad + \int E[g_{1,\tau|T=1}(Y)|X=x]p(x)s_x(x)f(x)\,dx\\ &\quad - \int E[g_{0,\tau|T=1}(Y)|X=x]p(x)s_x(x)f(x)\,dx.\end{aligned}$$

The efficient influence function for this case is equal to

$$\begin{aligned}&\psi_{\tau|T=1}(Y,T,X)\\ &\quad = \frac{T(g_{1,\tau|T=1}(Y)-E[g_{1,\tau|T=1}(Y)|X,T=1])}{p}\\ &\qquad - \frac{(1-T)p(X)(g_{0,\tau|T=1}(Y)-E[g_{0,\tau|T=1}(Y)|X,T=0])}{p(1-p(X))}\\ &\qquad + \frac{T-p(X)}{p}\\ &\qquad\quad \times \big(E[g_{1,\tau|T=1}(Y)|X,T=1]-E[g_{0,\tau|T=1}(Y)|X,T=0]\big)\\ &\qquad + \frac{p(X)}{p}\big(E[g_{1,\tau|T=1}(Y)|X,T=1]-E[g_{0,\tau|T=1}(Y)|X,T=0]\big).\end{aligned}$$

Whereas this influence function is in the set of scores, its expected value is zero and its variance is equal to $E[\psi^2_{\tau_{|T=1}}(Y,T,X)]$, which is the semiparametric efficiency bound for $\Delta_{\tau|T=1}$, $V_{\tau_{|T=1}}$. *Q.E.D.*

Dept. de Economia, Pontifícia Universidade Católica do Rio de Janeiro, Rua Marquês de São Vicente, 225-Gávea, 22453 Rio de Janeiro, RJ, Brazil; firpo@econ.puc-rio.br, http://www.econ.puc-rio.br/firpo/homepage_puc.htm.

Manuscript received August, 2004; final revision received June, 2006.

REFERENCES

BICKEL, P., C. A. J. KLAASSEN, Y. RITOV, AND J. A. WELLNER (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press. [34]

DEHEJIA, R., AND S. WAHBA (1999): "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053–1062. [1,3,7]

DINARDO, J., N. FORTIN, AND T. LEMIEUX (1996): "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach," *Econometrica*, 64, 1001–1044. [14]

HAHN, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66, 315–331. [34]

HALL, P. (1987): "On Kullback–Leibler Loss and Density Estimation," *The Annals of Statistics*, 15, 1491–1519. [3]

HECKMAN, J., AND J. HOTZ (1989): "Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," *Journal of the American Statistical Association*, 84, 862–874. [1]

HECKMAN, J., R. LALONDE, AND J. SMITH (1999): "The Economics and Econometrics of Active Labor Market Programs," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card. Amsterdam, 1277–1366. [3]

HIRANO, K., G. IMBENS, AND G. RIDDER (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189. [14,16]

HJÖRT, N., AND D. POLLARD (1993): "Asymptotics for Minimizers of Convex Processes," typescript, www.stat.yale.edu/~pollard/Papers/convex.pdf. [21-23]

IMBENS, G. (2003): "Sensitivity to Exogeneity Assumptions in Program Evaluation," *The American Economic Review*, 93, 126–132. [1]

LALONDE, R. J. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *The American Economic Review*, 76, 604–620. [1]

NEWEY, W. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99–135. [34]

——— (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382. [34]

——— (1995): "Convergence Rates for Series Estimators," in *Statistical Methods of Economics and Quantitative Economics: Essays in Honor of C. R. Rao*, ed. by G. S. Maddala, P. C. B. Phillips, and T. N. Srinavasan. Cambridge, MA: Blackwell, 254–275. [15]

——— (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168. [15]

ROSENBAUM, P., AND D. RUBIN (1984): "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score," *Journal of the American Statistical Association*, 79, 516–574. [3]

SMITH, J. A., AND P. E. TODD (2001): "Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods," *The American Economic Review*, 91, 112–118. [1,3]

——— (2005): "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*, 125, 305–353. [1,3]

WAND, M., AND M. JONES (1995): *Kernel Smoothing*. London: Chapman & Hall. [24,27]