Supplemental material for "Uniform Inference on Quantile Effects under Sharp Regression Discontinuity Designs"

Zhongjun Qu

and

Jungmo Yoon

This appendix is structured as follows. Section S.1 explains the bandwidth selectors used in the paper. Section S.2 discusses the hypothesis tests and the confidence bands related to $\delta^d(\tau)$ introduced in Section 2. Section S.3 develops Wald tests for the three hypotheses, assuming the second-order derivative of the conditional quantile function is continuous at the cut-off. Section S.4 provides a local asymptotic power analysis for the score and Wald tests. Section S.5 includes the proofs of the results given in the paper. Section S.6 reports additional simulation results, followed by several tables.

S.1 Bandwidth selection

This section discusses how the five selectors determine $h_{n,0.5}$, the bandwidth at the median. Then, bandwidths at other quantiles are computed using the link function of Yu and Jones (1998):

$$h_{n,\tau} = \left\{2\tau(1-\tau)/[\pi\phi(\Phi^{-1}(\tau))^2]\right\}^{1/5} h_{n,0.5},$$

where $\phi(\cdot)$ and $\Phi^{-1}(\cdot)$ are the density and quantile functions of a standard normal distribution.

The first two bandwidth selectors are based on the leave-one-out cross validation. They are simple modifications of the methods given in Ludwig and Miller (2007) and Imbens and Lemieux (2008), originally designed for the average treatment effect. Specifically, for a given candidate bandwidth h, we estimate the conditional median at x_i using the local linear regression, while leaving out (x_i, y_i), and denote the estimate by $\hat{Q}_h(0.5|x_i)$. Then, we compute $CV(h) = k^{-1}\sum_{i=1}^{k} |y_i - \hat{Q}_h(0.5|x_i)|$ and determine the bandwidth as $h_{n,0.5} = \arg\min_h CV(h)$. Because the focus here is on the responses near x_0, observations far from x_0 are less relevant. Therefore, following Imbens and Lemieux (2008), we use only half the observations that are closest to x_0 as evaluation points. These two selectors differ in terms of whether x_0 is treated as an interior or a boundary point. The first selector treats x_0 as an interior point, that is, utilizing observations on both sides of x_i when estimating the conditional median at x_i. This can be viewed as selecting the bandwidth by imposing the null hypothesis of no treatment effects. We denote the chosen bandwidth as $h_{n,\tau}^{cvi}$. The second selector treats x_0 as a boundary point. For example, if $x_i < x_0$, then only observations to the left of x_i are used when estimating the conditional median at x_i. This can be viewed as selecting the bandwidth under the alternative hypothesis. We denote the chosen bandwidth by $h_{n,\tau}^{cv}$.

The third bandwidth selector uses the minimum MSE bandwidth formula of Qu and Yoon (2015), while treating x_0 as an interior point. This leads to

$$h_{n,0.5} = \left(\frac{\int_{-\infty}^{\infty} K(v)^2 \, dv}{4\mu_2^2 f_X(x_0) f_{Y|X}(0.5|x_0)^2 \left(\frac{\partial^2 Q(0.5|x_0)}{\partial x^2}\right)^2} \right)^{1/5} n^{-1/5}. \tag{S.1}$$

The densities $f_X(x_0)$ and $f_{Y|X}(0.5|x_0)$ are estimated as follows. The marginal density estimate is $\hat{f}_X(x_0) = (nh_x)^{-1}\sum_{i=1}^{n} K((x_i - x_0)/h_x)$, where h_x is a bandwidth parameter. The conditional density estimate is

$$\hat{f}_{Y|X}(z|x_0) = \int \frac{1}{h_{yx}} K((z-y)/h_{yx}) \, d\hat{F}(y|x_0), \tag{S.2}$$

where h_{yx} is another bandwidth parameter, and $\hat{F}(y|x_0) = \sup\{\tau \in (0,1) | \hat{Q}(\tau|x_0) \leq y\}$ is the inverse function of $\hat{Q}(\tau|x_0)$, which is the estimated conditional quantile with the cross validation bandwidth. To implement (S.2), we draw samples from $\hat{F}(y|x_0)$ and apply the kernel density estimator to the sample with kernel $K(\cdot)$ and bandwidth h_{yx}. In practice, the bandwidth h_{yx} is

set to $2\widetilde{h}_{yx}$, with $\widetilde{h}_{yx}$ and h_x being the bandwidths determined using Silverman's rule of thumb formula. Finally, the second-order derivative $\partial^2 Q(0.5|x_0)/\partial x^2$ is estimated using the local cubic median regression. Its bandwidth will be set to 1.0 throughout the simulations (note that, in this case, the support of x is $[-1,1]$). We denote the resulting bandwidth by $h_{n,\tau}^{int}$.

The fourth bandwidth selector also uses the formula of Qu and Yoon (2015), but treats x_0 as a boundary point. This leads to the following bandwidth for $Q(0.5|x_0^+)$:

$$h_{n,0.5}^{+}=\left(\frac{\iota_1' N^{-1} M N^{-1} \iota_1}{4 f_X\left(x_0\right) f_{Y|X}\left(0.5|x_0^+\right)^2\left(\frac{\partial^2 Q(0.5|x_0^+)}{\partial x^2}\right)^2\left(\iota_1' N^{-1} L\right)^2}\right)^{1/5} n^{-1/5}, \tag{S.3}$$

where $\iota_1=(1\ \ 0)'$, N and M are 2-by-2 matrices with the $(i,j)th$ elements given by $\int_0^\infty u^{i+j-2}K\left(u\right)du$ and $\int_0^\infty u^{(i+j-2)}K(u)^2du$ and $L=[\ \int_0^\infty uK\left(u\right)du\ \ \int_0^\infty u^2K\left(u\right)du]'$. In the implementation, the derivative $\partial^2 Q(0.5|x_0^+)/\partial x^2$ is estimated in the same way as for the third bandwidth selector, but now uses only observations on the right side of x_0. The MSE optimal bandwidth for estimating $Q(0.5|x_0^-)$ satisfies the same expression as (S.3), but with $\int_{-\infty}^0$ replacing $\int_0^\infty$ and x_0^- replacing x_0^+. The one-sided conditional density $f_{Y|X}\left(0.5|x_0^+\right)$ uses the same formula as in (S.2), except that $\hat{F}(y|x_0)$ is replaced by $\hat{F}(y|x_0^+)$, which is computed by inverting $\hat{Q}(\tau|x_0^+)$. Finally, after obtaining estimates for $h_{n,0.5}^+$ and $h_{n,0.5}^-$, we use the smaller of the two to implement the tests. The motivation is that using a smaller bandwidth, although sacrificing some efficiency, will not erroneously introduce a large bias. We denote the bandwidth by $h_{n,\tau}^{bdy}$.

The fifth bandwidth selector is an adaptation of the Imbens and Kalyanaraman (2012) selector from the conditional mean to the conditional quantile setting. Instead of minimizing the MSEs associated with the conditional mean functions, Imbens and Kalyanaraman (2012) suggested minimizing the MSE associated with estimating their difference. For quantile treatment effects, calculations lead to the following bandwidth formula:

$$h_{n,0.5}=\left(\frac{\iota_1' N^{-1} M N^{-1} \iota_1\left(\frac{1}{f_{Y|X}\left(0.5|x_0^+\right)^2}+\frac{1}{f_{Y|X}\left(0.5|x_0^-\right)^2}\right)}{4\left(\iota_1' N^{-1} L\right)^2 f_X\left(x_0\right)\left(\left(\frac{\partial^2 Q(0.5|x_0^+)}{\partial x^2}-\frac{\partial^2 Q(0.5|x_0^-)}{\partial x^2}\right)^2+\left(r_-+r_+\right)\right)}\right)^{1/5} n^{-1/5}, \tag{S.4}$$

where r_+ and r_- are regularization terms that equal three times the variances of $\partial^2\hat{Q}(0.5|x_0^+)/\partial x^2$ and $\partial^2\hat{Q}(0.5|x_0^-)/\partial x^2$, respectively. Their purpose is to stabilize the bandwidth in situations where the second-order derivatives do not change at x_0, or when they are imprecisely estimated. The quantities r_- and r_+ depend on the following three factors for obtaining $\partial^2\hat{Q}(0.5|x_0^+)/\partial x^2$ and $\partial^2\hat{Q}(0.5|x_0^-)/\partial x^2$: the order of the local regressions, the kernel used, and the bandwidths. In simulations, we consider local quadratic regressions, the Epanechnikov kernel, and the bandwidth

$h_r = 0.5$. This leads to

$$r^+ = \frac{3}{nh_r^5} \frac{\iota_3' \bar{N}^{-1} \bar{M} \bar{N}^{-1} \iota_3'}{f_{Y|X}\left(0.5|x_0^+\right)^2 f_X\left(x_0\right)}, \tag{S.5}$$

where $\iota_3 = (0 \quad 0 \quad 1)'$, and $\bar{N}$ and $\bar{M}$ are 3-by-3 matrices, with the (i,j)-th elements given by $\int_0^\infty u^{i+j-2} K\left(u\right) du$ and $\int_0^\infty u^{i+j-2} K(u)^2 du$. The expression of r^- is the same as (S.5), but with $\int_{-\infty}^0$ and x_0^- replacing $\int_0^\infty$ and x_0^+, respectively. In the implementation, we rewrite (S.5) as

$$\frac{3}{nh_r^5} \frac{\iota_3' \left(f_X\left(x_0\right) \bar{N}\right)^{-1} \left[f_X\left(x_0\right) \bar{M}\right] \left(f_X\left(x_0\right) \bar{N}\right)^{-1} \iota_3'}{f_{Y|X}\left(0.5|x_0^+\right)^2}.$$

Then, the relevant quantities can be estimated using $(nh_r)^{-1} \sum_{i=1}^n \bar{z}_{i,\tau} \bar{z}_{i,\tau}' d_i K_{i,\tau} \to^p f_X\left(x_0\right) \bar{N}$ and $(nh_r)^{-1} \sum_{i=1}^n \bar{z}_{i,\tau} \bar{z}_{i,\tau}' d_i K_{i,\tau}^2 \to^p f_X\left(x_0\right) \bar{M}$. We denote the bandwidth by $h_{n,\tau}^{ik}$. We also experiment with estimating $\partial^2 Q(0.5|x_0^+)/\partial x^2$ and $\partial^2 Q(0.5|x_0^-)/\partial x^2$ using local cubic rather than quadratic regressions. Then, $(r_- + r_+)$ tends to take on substantially higher values than when using the local quadratic regression, often dominating the term $\left[\partial^2 Q(0.5|x_0^+)/\partial x^2 - \partial^2 Q(0.5|x_0^-)/\partial x^2\right]^2$. For this reason, we choose to use the quadratic regressions in the simulations and the empirical application.

Among the five selections, $h_{n,\tau}^{cvi}$ and $h_{n,\tau}^{int}$ are consistent with the principle of the score test because they impose the null hypothesis of no treatment effects. In addition, $h_{n,\tau}^{cv}$, $h_{n,\tau}^{bdy}$, and $h_{n,\tau}^{ik}$ are consistent with the principle of the Wald test. We use these pairings in the experimentations.

Finally, when implementing the tests with the bias estimation, we need additional bandwidth parameters for the regressions in (12). Motivated by the results in Calonico, Cattaneo, and Titiunik (2014), we let these equal the bandwidths for the local linear regressions (i.e., $b_{n,\tau} = h_{n,\tau}$ for all $\tau \in \mathcal{T}$) throughout the experimentations.

S.2 Hypothesis tests and confidence bands related to $\delta^d(\tau)$

This subsection shows how to test the hypotheses and to construct uniform confidence bands for $\delta^d(\tau)$. For any of the three specifications of $\delta^d(\tau)$ in Section 2, let $\delta_1(\tau)$ denote the quantity inside the first parentheses and $\delta_2(\tau)$ be the quantity inside the second parentheses. Then,

$$\delta^d(\tau) = \delta_1(\tau) - \delta_2(\tau).$$

The three null hypotheses of interest are: (i) $H_0^1 : \delta^d(\tau) = 0$ for any $\tau \in \mathcal{T}$; (ii) $H_0^2 : \delta^d(\tau) = c$ for some $c \in R$ for all $\tau \in \mathcal{T}$; and (iii) $H_0^3 : \delta^d(\tau) \geq 0$ for all $\tau \in \mathcal{T}$. Let n_1 and n_2 be the sample sizes, and $h_{n_1,\tau}$ and $h_{n_2,\tau}$ be the bandwidths when estimating $\delta_1(\tau)$ and $\delta_2(\tau)$. Let $f_{Y|X,j}(\tau|x_0^+)$, $f_{Y|X,j}(\tau|x_0^-)$, $d_{\tau,j}^+$, and $d_{\tau,j}^-$ $(j = 1, 2)$ be the respective conditional densities and biases. (In the case with two cut-offs, interpret $f_{Y|X,1}(\tau|x_0^+)$ as $f_{Y|X}(\tau|x_0^+)$ and $f_{Y|X,2}(\tau|x_0^+)$ as $f_{Y|X}(\tau|x_1^+)$.) Define $f_{Y|X}(\tau|x_0) = (f_{Y|X,1}(\tau|x_0^+) + f_{Y|X,1}(\tau|x_0^-) + f_{Y|X,2}(\tau|x_0^+) + f_{Y|X,2}(\tau|x_0^-))/4$.

S.2.1 Testing hypotheses assuming continuous second-order derivatives at the cutoffs

Define

$$W_n^d(\tau) = \sqrt{n_1 h_{n_1,\tau}}\hat{f}_{Y|X}(\tau|x_0)\left(\hat{\delta}_1(\tau) - \hat{\delta}_2(\tau)\right),$$

and consider the following test statistics.

$$\begin{aligned}
\text{For } H_0^1 \quad &: \quad WS_n^d(\mathcal{T}) = \sup_{\tau\in\mathcal{T}}\left|W_n^d(\tau)\right|, \\
\text{For } H_0^2 \quad &: \quad WH_n^d(\mathcal{T}) = \sup_{\tau\in\mathcal{T}}\left|W_n^d(\tau) - \frac{\sqrt{n_1 h_{n_1,\tau}}\hat{f}_{Y|X}(\tau|x_0)}{\int_{s\in\mathcal{T}}\sqrt{n_1 h_{n_1,s}}\hat{f}_{Y|X}(s|x_0)ds}\int_{\tau\in\mathcal{T}} W_n^d(\tau)d\tau\right|, \\
\text{For } H_0^3 \quad &: \quad WA_n^d(\mathcal{T}) = \sup_{\tau\in\mathcal{T}}\left|1\left(W_n^d(\tau)\le 0\right)W_n^d(\tau)\right|.
\end{aligned}$$

To present the limiting distributions of the test statistics, let $G_*^j(\tau)$ $(j=1,2)$ be two mutually independent Gaussian processes that are the limits of

$$\frac{1}{f_{X,j}(x_0)\sqrt{n_j h_{n_j,\tau}}}\sum_{i=1}^{n}\left(\tau - 1\left(u_i^0(\tau)\le 0\right)\right)\left\{\frac{f_{Y|X}(\tau|x_0)}{f_{Y|X,j}(\tau|x_0^+)}\Xi_{i,\tau,j}^+ d_i - \frac{f_{Y|X}(\tau|x_0)}{f_{Y|X,j}(\tau|x_0^-)}\Xi_{i,\tau,j}^-(1-d_i)\right\}K_{i,\tau,j},$$

where $\Xi_{i,\tau,j}^+$, $\Xi_{i,\tau,j}^-$, and $K_{i,\tau,j}$ are computed with bandwidth $h_{n_j,\tau}$. Let $\kappa(\tau)$ be the quantity defined in the Proposition below and $G_*^d(\tau) = G_*^1(\tau) - \kappa(\tau)G_*^2(\tau)$.

Proposition 3 *Assume the conditions in Lemma 2 hold for j=1,2 with* $\partial^2 Q(\tau|x_0^+)/\partial x^2 = \partial^2 Q(\tau|x_0^-)/\partial x^2$ *for all* $\tau\in\mathcal{T}$. *Assume* $\sqrt{n_1 h_{n_1,\tau}}/\sqrt{n_2 h_{n_2,\tau}} \to \kappa(\tau) > 0$. *Then:*

1. *Under* $\delta^d(\tau) = 0$ *for all* $\tau\in\mathcal{T}$, $WS_n^d(\mathcal{T}) \Rightarrow \sup_{\tau\in\mathcal{T}}\left|G_*^d(\tau)\right|$.

2. *Under* $\delta^d(\tau) = \delta$ *for all* $\tau\in\mathcal{T}$ *for some* $\delta\in\mathbb{R}$,

$$WH_n^d(\mathcal{T}) \Rightarrow \sup_{\tau\in\mathcal{T}}\left|G_*^d(\tau) - \frac{\sqrt{n_1 h_{n_1,\tau}}f_{Y|X}(\tau|x_0)}{\int_{s\in\mathcal{T}}\sqrt{n_1 h_{n_1,s}}f_{Y|X}(s|x_0)ds}\int_{\tau} G_*^d(\tau)d\tau\right|.$$

3. *Under the least favorable null hypothesis of* $\delta^d(\tau) = 0$ *for all* $\tau\in\mathcal{T}$,

$$WA_n^d(\mathcal{T}) \Rightarrow \sup_{\tau\in\mathcal{T}}\left|1\left(G_*^d(\tau)\le 0\right)G_*^d(\tau)\right|.$$

S.2.2 Testing hypotheses allowing discontinuous second-order derivatives at the cutoffs

Define

$$W_n^{R,d}(\tau) = \sqrt{n_1 h_{n_1,\tau}}\hat{f}_{Y|X}(\tau|x_0)\left(\hat{\delta}_1(\tau) - \hat{\delta}_2(\tau) - h_{n_1,\tau}^2(\hat{d}_{\tau,1}^+ - \hat{d}_{\tau,1}^-) + h_{n_2,\tau}^2(\hat{d}_{\tau,2}^+ - \hat{d}_{\tau,2}^-)\right),$$

where $\hat{d}^+_{\tau,j}$ and $\hat{d}^-_{\tau,j}$ are estimated with local quadratic regressions with bandwidth $h_{n_j,\tau}$ $(j = 1, 2)$. The tests are:

$$\begin{aligned}
\text{For } H_0^1 \quad &: \quad WS_n^{R,d}(\mathcal{T}) = \sup_{\tau\in\mathcal{T}} \left| W_n^{R,d}(\tau) \right|, \\
\text{For } H_0^2 \quad &: \quad WH_n^{R,d}(\mathcal{T}) = \sup_{\tau\in\mathcal{T}} \left| W_n^{R,d}(\tau) - \frac{\sqrt{n_1 h_{n_1,\tau}}\hat{f}_{Y|X}(\tau|x_0)}{\int_{s\in\mathcal{T}} \sqrt{n_1 h_{n_1,s}}\hat{f}_{Y|X}(s|x_0)ds} \int_{\tau\in\mathcal{T}} W_n^{R,d}(\tau)d\tau \right|, \\
\text{For } H_0^3 \quad &: \quad WA_n^{R,d}(\mathcal{T}) = \sup_{\tau\in\mathcal{T}} \left| 1\left(W_n^{R,d}(\tau) \leq 0\right) W_n^{R,d}(\tau) \right|.
\end{aligned}$$

Let $G_*^{R,j}(\tau)$ (j=1,2) be two independent copies of $G_*^R(\tau)$; see (15) in Section 5.2. Note that in $G_*^{R,j}(\tau)$, the bandwidth is equal to $h_{n_j,\tau}$. Define $G_*^{R,d}(\tau) = G_*^{R,1}(\tau) - \kappa(\tau)G_*^{R,2}(\tau)$.

Proposition 4 *Let the conditions in Lemma 2 and Lemma 3 hold for j=1,2. Assume* $\sqrt{n_1 h_{n_1,\tau}}/\sqrt{n_2 h_{n_2,\tau}} \to \kappa(\tau) > 0$. *Then:*

1. *Under* $\delta(\tau) = 0$ *for all* $\tau \in \mathcal{T}$, $WS_n^{R,d}(\mathcal{T}) - \sup_{\tau\in\mathcal{T}} \left| G_*^{R,d}(\tau) \right| = o_p(1)$.

2. *Under* $\delta(\tau) = \delta$ *for all* $\tau \in \mathcal{T}$ *for some* $\delta \in \mathbb{R}$,

$$WH_n^{R,d}(\mathcal{T}) - \sup_{\tau\in\mathcal{T}} \left| G_*^{R,d}(\tau) - \frac{\sqrt{nh_{n,\tau}}\hat{f}_{Y|X}(\tau|x_0)}{\int_{s\in\mathcal{T}} \sqrt{nh_{n,s}}\hat{f}_{Y|X}(s|x_0)ds} \int_{\tau\in\mathcal{T}} G_*^{R,d}(\tau)d\tau \right| = o_p(1).$$

3. *Under the least favorable null hypothesis of* $\delta(\tau) = 0$ *for all* $\tau \in \mathcal{T}$,

$$WA_n^{R,d}(\mathcal{T}) - \sup_{\tau\in\mathcal{T}} \left| 1\left(G_*^{R,d}(\tau) \leq 0\right) G_*^{R,d}(\tau) \right| = o_p(1).$$

The relevant critical values can be obtained using simulations.

S.2.3 Uniform confidence bands for $\delta^d(\tau)$

A uniform band can be obtained by inverting the Wald tests for the hypothesis H_0^1. In the case with continuous second-order derivatives, let c_p^d be the $(1-p)$ percentile of the distribution of $\sup_{\tau\in\mathcal{T}} \left|G_*^d(\tau)\right|$. The confidence band for $\delta^d(\tau)$ is then given by

$$\hat{\delta}_1(\tau) - \hat{\delta}_2(\tau) \pm \frac{c_p^d}{\sqrt{n_1 h_{n_1,\tau}}\hat{f}_{Y|X}(\tau|x_0)}.$$

When discontinuous second-order derivatives are allowed, let $c_p^{R,d}$ be the $(1-p)$ percentile of the distribution of $\sup_{\tau\in\mathcal{T}} \left|G_*^{R,d}(\tau)\right|$. The uniform band is given by

$$\hat{\delta}_1(\tau) - \hat{\delta}_2(\tau) - h_{n_1,\tau}^2(\hat{d}^+_{\tau,1} - \hat{d}^-_{\tau,1}) + h_{n_2,\tau}^2(\hat{d}^+_{\tau,2} - \hat{d}^-_{\tau,2}) \pm \frac{c_p^{R,d}}{\sqrt{n_1 h_{n_1,\tau}}\hat{f}_{Y|X}(\tau|x_0)}.$$

S.3 Wald tests assuming $\partial^2 Q(\tau|x_0^+)/\partial x^2 = \partial^2 Q(\tau|x_0^-)/\partial x^2$

Define

$$W_n(\tau) = \sqrt{nh_{n,\tau}}\hat{f}_{Y|X}(\tau|x_0)\hat{\delta}(\tau), \tag{S.6}$$

Treatment significance. This hypothesis can be tested using a Kolmogorov−Smirnov-type test:

$$WS_n(\mathcal{T}) = \sup_{\tau\in\mathcal{T}} |W_n(\tau)| .$$

Treatment homogeneity. This hypothesis can be tested by measuring the deviation of $W_n(\tau)$ from the average of $W_n(\tau)$ over $\mathcal{T}$:

$$WH_n(\mathcal{T}) = \sup_{\tau\in\mathcal{T}} \left| W_n(\tau) - \frac{\sqrt{nh_{n,\tau}}\hat{f}_{Y|X}(\tau|x_0)}{\int_{s\in\mathcal{T}} \sqrt{nh_{n,s}}\hat{f}_{Y|X}(s|x_0)ds} \int_{\tau\in\mathcal{T}} W_n(\tau)d\tau \right| .$$

Treatment unambiguity. To test this hypothesis, we determine whether the treatment can be detrimental at some unknown quantiles, using

$$WA_n(\mathcal{T}) = \sup_{\tau\in\mathcal{T}} |1(W_n(\tau) \le 0) W_n(\tau)| .$$

Let $G_1(\tau)$ be a zero-mean continuous Gaussian process with a covariance function that satisfies

$$E[G_1(t)G_1(s)] = \frac{(t\wedge s - ts)}{f_X(x_0)(\mu_0^+\mu_2^+ - (\mu_1^+)^2)^2 (\kappa(t)\kappa(s))^{1/2}} \int_{-\infty}^{\infty} H(t)H(s)K\left(\frac{u}{\kappa(t)}\right) K\left(\frac{u}{\kappa(s)}\right) du, \tag{S.7}$$

where

$$H(\tau) = \frac{f_{Y|X}(\tau|x_0)}{f_{Y|X}(\tau|x_0^+)} \left(\mu_2^+ - \left(\frac{u}{\kappa(\tau)}\right)\mu_1^+\right) I(u\ge 0) - \frac{f_{Y|X}(\tau|x_0)}{f_{Y|X}(\tau|x_0^-)} \left(\mu_2^- - \left(\frac{u}{\kappa(\tau)}\right)\mu_1^-\right)(1 - I(u\ge 0)) .$$

Proposition 5 *Assume the same conditions as in Lemma 2 hold, with* $\partial^2 Q(\tau|x_0^+)/\partial x^2 = \partial^2 Q(\tau|x_0^-)/\partial x^2$ *for all* $\tau\in\mathcal{T}$*. Then:*

1. *Under* $\delta(\tau) = 0$ *for all* $\tau\in\mathcal{T}$, $WS_n(\mathcal{T}) \Rightarrow \sup_{\tau\in\mathcal{T}} |G_1(\tau)|$.
2. *Under* $\delta(\tau) = \delta$ *for all* $\tau\in\mathcal{T}$ *for some* $\delta\in\mathbb{R}$,

$$WH_n(\mathcal{T}) \Rightarrow \sup_{\tau\in\mathcal{T}} \left| G_1(\tau) - \frac{\sqrt{nh_{n,\tau}}f_{Y|X}(\tau|x_0)}{\int_{s\in\mathcal{T}} \sqrt{nh_{n,s}}f_{Y|X}(s|x_0)ds} \int_{\tau} G_1(\tau)d\tau \right| .$$

3. *Under the least favorable null hypothesis of* $\delta(\tau) = 0$ *for all* $\tau\in\mathcal{T}$ *(this is explained in the proof),*

$$WA_n(\mathcal{T}) \Rightarrow \sup_{\tau\in\mathcal{T}} |1(G_1(\tau)\le 0) G_1(\tau)| .$$

Proof of Proposition 5. In all three results, the effects are homogeneous across quantiles. This implies $f_{Y|X}(\tau|x_0^+) = f_{Y|X}(\tau|x_0^-)$ and, consequently,

$$W_{n,c}(\tau) = \frac{1}{f_X(x_0)\sqrt{nh_{n,\tau}}} \sum_{i=1}^{n} \left(\tau - I\left(u_i^0(\tau) \leq 0\right)\right) \frac{(2d_i - 1)\mu_2^+ - \left(\frac{x_i - x_0}{h_\tau}\right)\mu_1^+}{\mu_0^+\mu_2^+ - \left(\mu_1^+\right)^2} K_{i,\tau} + o_p(1).$$

The results then follow from the same arguments as in the proof of Proposition 1. For Case 3, the reason why "$\delta(\tau) = 0$ for all $\tau \in \mathcal{T}$ " is the least favorable null for the treatment unambiguity hypothesis is as follows. Define $M(\tau) = \sqrt{nh_{n,\tau}}\hat{f}_{Y|X}(\tau|x_0)\delta(\tau)$. Then, for any $\delta(\tau)$ satisfying the null hypothesis (i.e., $\delta(\tau) \geq 0$ for all $\tau \in \mathcal{T}$), the following two inequalities always hold because $M(\tau) \geq 0$:

$$\begin{aligned} |1\left(W_n(\tau) \leq 0\right) W_n(\tau)| &\leq |1\left(W_n(\tau) \leq 0\right)\left(W_n(\tau) - M(\tau)\right)| \\ &\leq |1\left(W_n(\tau) - M(\tau) \leq 0\right)\left(W_n(\tau) - M(\tau)\right)|. \end{aligned} \tag{S.8}$$

The term $W_n(\tau) - M(\tau)$ is equal to $W_{n,c}(\tau)$, defined in (14). Therefore, it satisfies the approximation given in Lemma 2 for any $\delta(\tau) \geq 0$. As a result, the supremum of (S.8) converges to $\sup_{\tau\in\mathcal{T}} |1\left(G_1(\tau) \leq 0\right) G_1(\tau)|$ under $\delta(\tau) \geq 0$. This shows that the test may be conservative if $\delta(\tau) \geq 0$ but $\delta(\tau)$ is not always zero. The test will not over-reject the null hypothesis. This completes the proof.

S.4 Local asymptotic power analysis

The local alternatives are specified as follows. When testing for the treatment significance and unambiguity hypotheses, let

$$Q(\tau|x_0^+) - Q(\tau|x_0^-) = (nh_n)^{-1/2}\eta(\tau), \tag{S.9}$$

with $|\eta(\tau)| < +\infty$ for all $\tau \in \mathcal{T}$. When testing for the treatment homogeneity hypothesis, let

$$Q(\tau|x_0^+) - Q(\tau|x_0^-) = \delta + (nh_n)^{-1/2}\eta(\tau), \tag{S.10}$$

with $|\delta| < +\infty$ and $|\eta(\tau)| < +\infty$ for all $\tau \in \mathcal{T}$. The bandwidth h_n satisfies Assumption (5). The quantities δ and $\eta(\tau)$ are fixed as $n \to \infty$.

Proposition 6 *Assume the same conditions as in Lemma 2 hold with $\partial^2 Q(\tau|x_0^+)/\partial x^2 = \partial^2 Q(\tau|x_0^-)/\partial x^2$ for all $\tau \in \mathcal{T}$. Let*

$$\widetilde{G}_1(\tau) = G_1(\tau) + c(\tau)^{1/2} f_{Y|X}(\tau|x_0)\eta(\tau),$$

where $c(\tau)$ is defined in Assumption (5). Then:

1. *Under (S.9),* $R_n(\mathcal{T}) \Rightarrow f_X(x_0)\left(\frac{\mu_0^+\mu_2^+-(\mu_1^+)^2}{2\mu_2^+}\right)\sup_{\tau\in\mathcal{T}}\left|\widetilde{G}_1(\tau)\right|$.

2. *Under (S.9),* $WS_n(\mathcal{T}) \Rightarrow \sup_{\tau\in\mathcal{T}}\left|\widetilde{G}_1(\tau)\right|$.

3. *Under (S.10),*

$$WH_n(\mathcal{T}) \Rightarrow \sup_{\tau\in\mathcal{T}}\left|\widetilde{G}_1(\tau) - \frac{\sqrt{nh_{n,\tau}}f_{Y|X}(\tau|x_0)}{\int_{s\in\mathcal{T}}\sqrt{nh_{n,s}}f_{Y|X}(s|x_0)ds}\int_\tau \widetilde{G}_1(\tau)d\tau\right|.$$

4. *Under (S.9) with* $\eta(\tau) < 0$ *for all* $\tau \in \mathcal{T}$,

$$WA_n(\mathcal{T}) \Rightarrow \sup_{\tau\in\mathcal{T}}\left|1\left(\widetilde{G}_1(\tau)\leq 0\right)\widetilde{G}_1(\tau)\right|.$$

The proof uses the same arguments as that of Lemmas 1 and 2. It is omitted. Interestingly, the first two results show that the score and Wald tests for the treatment significance hypothesis have the same local asymptotic power against the sequence (S.9). This follows after noting that, under the null hypothesis, their covariance functions satisfy

$$E(G(t)G(s)) = f_X(x_0)^2\left(\frac{\mu_0^+\mu_2^+ - (\mu_1^+)^2}{2\mu_2^+}\right)^2 E(G_1(t)G_1(s)).$$

In addition, the four results show that the tests can have nontrivial power against alternatives of order $(nh_n)^{-1/2}$. Finally, what matters for power is not only the difference $Q(\tau|x_0^+) - Q(\tau|x_0^-)$, but also the conditional density and the bandwidth. Everything else being equal, the power is higher if the departure from the null occurs in a dense region or at a place where the bandwidth is wider.

S.5 Proofs of results in the paper

Proof of Lemma 1. For any $\alpha(\tau) \in \mathbb{R}$ and $\beta(\tau) \in \mathbb{R}$, define

$$\begin{aligned} e_i(\tau) &= Q(\tau|x_0) + (x_i - x_0)'\frac{\partial Q(\tau|x_0)}{\partial x} - Q(\tau|x_i), \\ \phi(\tau) &= \sqrt{nh_{n,\tau}}\begin{pmatrix} \alpha(\tau) - Q(\tau|x_0) \\ h_{n,\tau}\left(\beta(\tau) - \frac{\partial Q(\tau|x_0)}{\partial x}\right)\end{pmatrix} \quad \text{and} \quad z_{i,\tau}' = \left(1, \frac{x_i - x_0}{h_{n,\tau}}\right). \end{aligned}$$

Applying (7), we can write

$$u_i(\tau) = u_i^0(\tau) - e_i(\tau) - (nh_{n,\tau})^{-1/2}z_{i,\tau}'\phi(\tau).$$

Consequently,

$$R_n(\tau) = (nh_{n,\tau})^{-1/2}\sum_{i=1}^n\left\{\tau - 1[u_i^0(\tau) \leq (nh_{n,\tau})^{-1/2}z_{i,\tau}'\phi(\tau) + e_i(\tau)]\right\}d_iK_{i,\tau}.$$

To establish the asymptotic property of $R_n(\tau)$, we need to analyze both the effect of the parameter estimation and that of the local linear approximation. To this end, define

$$\begin{aligned} S_n\left(\tau, \phi(\tau), e_i\left(\tau\right)\right) &= (nh_{n,\tau})^{-1/2} \sum_{i=1}^{n} \left\{ P\left[u_i^0(\tau) \leq (nh_{n,\tau})^{-1/2} z_{i,\tau}' \phi(\tau) + e_i(\tau) \middle| x_i \right] \right. \\ &\quad \left. - 1\left[u_i^0(\tau) \leq (nh_{n,\tau})^{-1/2} z_{i,\tau}' \phi(\tau) + e_i(\tau) \right] \right\} d_i K_{i,\tau}. \end{aligned}$$

Let $\hat{\phi}(\tau)$ and $S_n(\tau, \hat{\phi}(\tau), e_i\left(\tau\right))$ equal $\phi(\tau)$ and $S_n\left(\tau, \phi(\tau), e_i\left(\tau\right)\right)$, but evaluated at $\hat{\alpha}(\tau)$ and $\hat{\beta}(\tau)$. Then, by adding and subtracting terms:

$$\begin{aligned} R_n(\tau) &= S_n\left(\tau, 0, 0\right) && \text{(Term 1)} \\ &\quad + \left\{ S_n\left(\tau, 0, e_i\left(\tau\right)\right) - S_n\left(\tau, 0, 0\right) \right\} && \text{(Term 2)} \\ &\quad + \left\{ S_n(\tau, \hat{\phi}(\tau), e_i\left(\tau\right)) - S_n\left(\tau, 0, e_i\left(\tau\right)\right) \right\} && \text{(Term 3)} \\ &\quad + (nh_{n,\tau})^{-1/2} \sum_{i=1}^{n} \left\{ \tau - P(\, u_i^0(\tau) \leq e_i\left(\tau\right) + (nh_{n,\tau})^{-1/2} z_{i,\tau}' \hat{\phi}\left(\tau\right) \middle| x_i) \right\} d_i K_{i,\tau} && \text{(Term 4).} \end{aligned}$$

Term 1 depends only on the data generating process. Term 2 depends on the remainder term from the local linear approximation. Terms 3 and 4 are affected by the parameter estimation. By Theorems 2 and 3 in Qu and Yoon (2015), the inequality constraints (or rearrangement) have no first-order effect on $\hat{\phi}(\tau)$. Therefore, we can treat $\hat{\phi}(\tau)$ as the estimator obtained by applying quantile-by-quantile local linear regressions without imposing any constraints (or rearrangement). Further, Qu and Yoon (2015, Step 1 in the proof of Theorem 1) show that $\Pr(\sup_{\tau \in \mathcal{T}} ||\widehat{\phi}(\tau)|| \leq \log^{1/2}(nh_{n,\tau})) \to 1$. Therefore, it suffices to consider the set $\{\phi(\tau) : ||\phi(\tau)|| \leq \log^{1/2}(nh_{n,\tau})\}$ when analyzing $R_n(\tau)$.

We study Terms 1 to 4 separately. By Lemma B.5 in Qu and Yoon (2015), $\sup_{\tau \in \mathcal{T}} ||(\text{Term } 2)|| = o_p\left(1\right)$ and $\sup_{\tau \in \mathcal{T}} ||(\text{Term } 3)|| = o_p\left(1\right)$.[1] Apply the mean value theorem:

$$\begin{aligned} (\text{Term } 4) &= -(nh_{n,\tau})^{-1/2} \sum_{i=1}^{n} f_{Y|X}(\widetilde{y}_i|x_i)\, e_i(\tau) d_i K_{i,\tau} - \left(\frac{1}{nh_{n,\tau}} \sum_{i=1}^{n} f_{Y|X}(\widetilde{y}_i|x_i) K_{i,\tau} d_i z_{i,\tau}' \right) \hat{\phi}(\tau) \\ &= A_{n,1}(\tau) + A_{n,2}(\tau) \hat{\phi}(\tau), \end{aligned}$$

where $\widetilde{y}_i$ lies between $Q(\tau|x_i)$ and $Q(\tau|x_i) + e_i(\tau) + (nh_{n,\tau}^d)^{-1/2} z_{i,\tau}' \hat{\phi}$. To analyze $A_{n,1}(\tau)$, note that

$$e_i(\tau) = -\frac{1}{2} h_{n,\tau}^2 \left(\frac{x_i - x_0}{h_{n,\tau}} \right)^2 \frac{\partial Q^2(\tau|x_0)}{\partial x^2} + o(h_{n,\tau}^2) \text{ uniformly over } \tau \in \mathcal{T}.$$

[1] Lemma B.5 focuses on Term 3 while establishing the order of Term 2 as an intermediate result; see the second term on the right-hand side of (B.8) on page 18.

Therefore, uniformly over $\tau \in \mathcal{T}$,

$$\begin{aligned} A_{n,1}(\tau) &= \frac{1}{2}(nh_{n,\tau}^5)^{1/2}\frac{\partial^2 Q(\tau|x_0)}{\partial x^2}\left\{\frac{1}{nh_{n,\tau}}\sum_{i=1}^n f_{Y|X}(\widetilde{y}_i|x_i)\left(\frac{x_i-x_0}{h_{n,\tau}}\right)^2 d_i K_{i,\tau}\right\}+o_p(1) \\ &= \frac{1}{2}(nh_{n,\tau}^5)^{1/2}\frac{\partial^2 Q(\tau|x_0)}{\partial x^2}f_{Y|X}(\tau|x_0)\left\{\frac{1}{nh_{n,\tau}}\sum_{i=1}^n \left(\frac{x_i-x_0}{h_{n,\tau}}\right)^2 d_i K_{i,\tau}\right\}+o_p(1) \\ &= \frac{1}{2}(nh_{n,\tau}^5)^{1/2}f_{Y|X}(\tau|x_0)f_X(x_0)\frac{\partial^2 Q(\tau|x_0)}{\partial x^2}\mu_2^+ + o_p(1), \end{aligned}$$

where the second equality holds because x_i is in a vanishing neighborhood of x_0, and the third equality is by the uniform law of large numbers. By similar arguments, $A_{n,2}(\tau) = -f_{Y|X}(\tau|x_0)f_X(x_0)(\mu_0^+ \mu_1^+) + o_p(1)$. Finally, for $\hat{\phi}(\tau)$, apply Theorem 1 of Qu and Yoon (2015, see (A4) on page 15):

$$\begin{aligned} \hat{\phi}(\tau) &= \frac{1}{f_{Y|X}(\tau|x_0)\,f_X(x_0)}\begin{pmatrix}1 & 0\\ 0 & \mu_2^{-1}\end{pmatrix}\left\{(nh_{n,\tau})^{-1/2}\sum_{i=1}^n\left(\tau - 1\left(u_i^0(\tau)\le 0\right)\right)z_{i,\tau}K_{i,\tau}\right. \\ &\quad \left. + \frac{1}{2}\sqrt{nh_{n,\tau}^5}f_{Y|X}(\tau|x_0)\,f_X(x_0)\frac{\partial^2 Q(\tau|x_0)}{\partial x^2}\begin{pmatrix}\mu_2\\ \mu_3\end{pmatrix}\right\}+o_p(1). \end{aligned}$$

The results for $A_{n,1}(\tau), A_{n,2}(\tau)$, and $\hat{\phi}(\tau)$ jointly imply, uniformly over $\tau \in \mathcal{T}$:

$$\begin{aligned} & A_{n,1}(\tau) + A_{n,2}(\tau)\hat{\phi}(\tau) \\ = \; & -(nh_{n,\tau})^{-1/2}\sum_{i=1}^n\left(\tau - 1(u_i^0(\tau)\le 0)\right)\left(\mu_0^+ + \frac{\mu_1^+}{\mu_2}\left(\frac{x_i-x_0}{h_{n,\tau}}\right)\right)K_{i,\tau} \\ - \; & \frac{1}{2}\sqrt{nh_{n,\tau}^5}f_{Y|X}(\tau|x_0)f_X(x_0)\frac{\partial Q^2(\tau|x_0)}{\partial x^2}\left(\mu_0^+\mu_2 + \frac{\mu_1^+\mu_3}{\mu_2} - \mu_2^+\right)+o_p(1). \end{aligned}$$

Combining the results for Terms 1 to 4, we have

$$\begin{aligned} R_n(\tau) &= (nh_{n,\tau})^{-1/2}\sum_{i=1}^n\left(\tau - 1(u_i^0(\tau)\le 0)\right)\left(d_i - \mu_0^+ - \frac{\mu_1^+}{\mu_2}\left(\frac{x_i-x_0}{h_{n,\tau}}\right)\right)K_{i,\tau} \\ &+ \frac{1}{2}\sqrt{nh_{n,\tau}^5}f_{Y|X}(\tau|x_0)f_X(x_0)\frac{\partial Q^2(\tau|x_0)}{\partial x^2}\left(\mu_2^+ - \mu_0^+\mu_2 - \frac{\mu_1^+\mu_3}{\mu_2}\right)+o_p(1). \end{aligned}$$

Because the kernel is symmetric, $\mu_3 = 0, \mu_0^+ = 1/2$ and $\mu_2^+ = 0.5\mu_2$. As a result, $\mu_2^+ - \mu_0^+\mu_2 - \frac{\mu_1^+\mu_3}{\mu_2} = 0$. This completes the proof.

Proof of Proposition 1. It suffices to consider the leading term on the right-hand side in Lemma 1. For any fixed $\tau \in \mathcal{T}$, this term satisfies the central limit theorem. Its stochastic equicontinuity with respect to τ is implied by Lemma B3 in Qu and Yoon (2015). The result follows because the supremum operator is continuous when taken over a compact set.

Proof of Lemma 2. By Theorem 1 in Qu and Yoon (2015),

$$\begin{aligned}
&\sqrt{nh_{n,\tau}}\left(\hat{Q}(\tau|x_0^+) - Q(\tau|x_0^+)\right) \qquad (S.11)\\
= \;& \sqrt{nh_{n,\tau}^5}d_\tau^+ + \frac{\iota_1' N^{-1}\left(nh_{n,\tau}\right)^{-1/2}\sum_{i=1}^n\left(\tau - 1(u_i^0(\tau)\le 0)\right)d_i z_{i,\tau}K_{i,\tau}}{f_X(x)\, f_{Y|X}\left(\tau|x_0^+\right)} + o_p(1),
\end{aligned}$$

where $\iota_1 = (1,0)', u\in\mathbb{R}, \bar{u} = (1,u')', d_\tau^+ = \frac{1}{2}\iota_1' N^{-1}\frac{\partial^2 Q(\tau|x_0^+)}{\partial x^2}\int_0^\infty u^2\bar{u}K(u)\,du$, and $N = \int_0^\infty \bar{u}\bar{u}'K(u)\,du$. Because $\iota_1' N^{-1} = (\mu_0^+\mu_2^+ - (\mu_1^+)^2)^{-1}[\mu_2^+ \quad -\mu_1^+]$, the first term on the right side of (S.11) is equal to

$$\frac{1}{2}\sqrt{nh_{n,\tau}^5}\frac{\partial Q^2(\tau|x_0^+)}{\partial x^2}\frac{(\mu_2^+)^2 - \mu_1^+\mu_3^+}{\mu_0^+\mu_2^+ - (\mu_1^+)^2}, \qquad (S.12)$$

while the second term is equal to

$$\frac{\left(nh_{n,\tau}\right)^{-1/2}\sum_{i=1}^n\left(\tau - 1(u_i^0(\tau)\le 0)\right)\Xi_{i,\tau}^+ d_i K_{i,\tau}}{f_X(x)\, f_{Y|X}\left(\tau|x_0^+\right)}. \qquad (S.13)$$

Applying the same arguments to $\hat{Q}(\tau|x_0^-)$, we have

$$\begin{aligned}
\sqrt{nh_{n,\tau}}\left(\hat{Q}(\tau|x_0^-) - Q(\tau|x_0^-)\right) &= \frac{1}{2}\sqrt{nh_{n,\tau}^5}\frac{\partial Q^2(\tau|x_0^-)}{\partial x^2}\frac{(\mu_2^-)^2 - \mu_1^-\mu_3^-}{\mu_0^-\mu_2^- - (\mu_1^-)^2} \qquad (S.14)\\
&\quad + \frac{\left(nh_{n,\tau}\right)^{-1/2}\sum_{i=1}^n\left(\tau - 1(u_i^0(\tau)\le 0)\right)\Xi_{i,\tau}^-(1-d_i)K_{i,\tau}}{f_X(x)\, f_{Y|X}\left(\tau|x_0^-\right)} + o_p(1).
\end{aligned}$$

Combining (S.12), (S.13) and (S.14) leads to the desired result.

Proof of Lemma 3. It suffices to show that $\sqrt{nb_{n,\tau}^5}\left(\hat{d}_\tau^+ - d_\tau^+\right) = D_2^+(\tau) + o_p(1)$ uniformly over $\tau\in\mathcal{T}$. The proof is similar to that of Lemma 1. To reflect this, we define the notation analogously. Let

$$\begin{aligned}
\bar{u}_i(\tau) &= y_i - \alpha^+(\tau) - (x_i - x)\beta^+(\tau) - (x_i - x)^2\gamma^+(\tau),\\
\bar{e}_i(\tau) &= \left[Q(\tau|x_0^+) + (x_i - x_0)\frac{\partial Q(\tau|x_0^+)}{\partial x} - (x_i - x_0)^2\frac{1}{2}\frac{\partial Q^2(\tau|x_0^+)}{\partial x^2}\right] - Q(\tau|x_i),\\
\bar{\phi}(\tau) &= \sqrt{nb_{n,\tau}}\begin{pmatrix} \alpha^+(\tau) - Q(\tau|x_0^+)\\ b_{n,\tau}(\beta^+(\tau) - \frac{\partial Q(\tau|x_0^+)}{\partial x})\\ b_{n,\tau}^2(\gamma^+(\tau) - \frac{1}{2}\frac{\partial Q^2(\tau|x_0^+)}{\partial x^2})\end{pmatrix}, \text{ and } \bar{z}_{i,\tau}' = \begin{bmatrix} 1\\ (x_i - x_0)/b_{n,\tau}\\ (x_i - x_0)^2/b_{n,\tau}^2\end{bmatrix}.
\end{aligned}$$

Define

$$\begin{aligned}
\bar{S}_n\left(\tau, \bar{\phi}(\tau), \bar{e}_i(\tau)\right) &= (nb_n)^{-1/2}\sum_{i=1}^n\left\{P\left[u_i^0(\tau) \le (nb_{n,\tau})^{-1/2}\bar{z}_{i,\tau}'\bar{\phi}(\tau) + \bar{e}_i(\tau)\middle| x_i\right]\right.\\
&\quad \left. - 1\left[u_i^0(\tau)\le (nb_{n,\tau})^{-1/2}\bar{z}_{i,\tau}'\bar{\phi}(\tau) + \bar{e}_i(\tau)\right]\right\} d_i\bar{K}_{i,\tau}.
\end{aligned}$$

Note that $\bar{e}_i(\tau)$ satisfies

$$\bar{e}_i(\tau) = -\frac{1}{3!}\left(\frac{x_i - x}{b_{n,\tau}}\right)^3 \frac{\partial^3 Q(\tau|x_0^+)}{\partial x^3} b_{n,\tau}^3 + o\left(b_{n,\tau}^3\right). \tag{S.15}$$

Let $\widehat{\bar{\phi}}(\tau)$ equal $\bar{\phi}(\tau)$, but with $\alpha^+(\tau), \beta^+(\tau)$, and $\gamma^+(\tau)$ replaced by the estimates from the local quadratic regression. Applying the subgradient condition, we have

$$(nb_n)^{-1/2}\sum_{i=1}^n 1\left[u_i^0(\tau) \le (nb_{n,\tau})^{-1/2}\bar{z}_{i,\tau}'\widehat{\bar{\phi}}(\tau) + e_i(\tau)\right]\Big\} d_i\bar{K}_{i,\tau} = o_p(1)$$

uniformly over $\tau \in \mathcal{T}$. This implies

$$\begin{aligned}
&\{\bar{S}_n(\tau, \widehat{\bar{\phi}}(\tau), \bar{e}_i(\tau)) - \bar{S}_n(\tau, 0, \bar{e}_i(\tau))\} \\
&+ \left\{\bar{S}_n(\tau, 0, \bar{e}_i(\tau)) - \bar{S}_n(\tau, 0, 0)\right\} + \bar{S}_n(\tau, 0, 0) \\
&+(nb_n)^{-1/2}\sum_{i=1}^n \left\{\tau - P(u_i^0(\tau) \le \bar{e}_i(\tau) + (nb_{n,\tau})^{-1/2}\bar{z}_{i,\tau}'\widehat{\bar{\phi}}(\tau)\Big| x_i)\right\} d_i\bar{z}_{i,\tau}\bar{K}_{i,\tau} = o_p(1).
\end{aligned} \tag{S.16}$$

The terms in the first two curly brackets are $o_p(1)$ uniformly. Applying a first-order Taylor expansion to the last term, we obtain:

$$-(nb_n)^{-1/2}\sum_{i=1}^n f_{Y|X}(\widetilde{y}_i|x_i)\,\bar{e}_i(\tau)\,d_i\bar{z}_{i,\tau}\bar{K}_{i,\tau} - (nb_n)^{-1/2}(nb_{n,\tau})^{-1/2}\left(\sum_{i=1}^n f_{Y|X}(\widetilde{y}_i|x_i)\,d_iK_{i,\tau}\bar{z}_{i,\tau}\bar{z}_{i,\tau}'\right)\widehat{\bar{\phi}}(\tau),$$

where $\widetilde{y}_i$ lies between $Q(\tau|x_i)$ and $Q(\tau|x_i) + e_i(\tau) + (nb_{n,\tau})^{-1/2}d_i\bar{z}_{i,\tau}'\widehat{\bar{\phi}}(\tau)$. As a result,

$$\begin{aligned}
\widehat{\bar{\phi}}(\tau) \;=\; & \left(f_{Y|X}\left(\tau|x_0^+\right) f_X(x_0)\,\bar{N}^+\right)^{-1} \\
& \left\{\left(\frac{b_n}{b_{n,\tau}}\right)^{1/2}\bar{S}_n(\tau, 0, 0) - (nb_{n,\tau})^{-1/2} f_{Y|X}\left(\tau|x_0^+\right)\sum_{i=1}^n \bar{e}_i(\tau)\,d_i\bar{z}_{i,\tau}\bar{K}_{i,\tau}\right\} + o_p(1).
\end{aligned}$$

The term involving $\bar{e}_i(\tau)$ is negligible because $nb_{n,\tau}^7 = o(1)$. Therefore,

$$\widehat{\bar{\phi}}(\tau) = \left(f_{Y|X}\left(\tau|x_0^+\right) f_X(x_0)\,\bar{N}^+\right)^{-1}(b_n/b_{n,\tau})^{1/2}\,\bar{S}_n(\tau, 0, 0) + o_p(1).$$

Multiplying both sides by $\Gamma\iota_3'$ leads to the desired result.

Proof of Proposition 2. By Lemma 2 and Lemma 3,

$$W_n^R(\tau) = G_*^R(\tau) + o_p(1)$$

uniformly over $\mathcal{T}$. Then, the proof can be completed by applying the same arguments as those in the proof of Proposition 1.

Proof of the validity of the procedure in Remark 2. The proof is given in four steps, using similar arguments as in Politis and Romano (1994) and Hahn (1995). It allows for two

possible bandwidth sequences: (i) $b_{n,\tau}/h_{n,\tau} \to \infty$ for all $\tau \in \mathcal{T}$. This corresponds to using a larger bandwidth for the local quadratic regression than the local linear regression. (ii) $b_{n,\tau}/h_{n,\tau} = r(\tau)$ with $0 < r(\tau) < \infty$ for all $\tau \in \mathcal{T}$. This corresponds to using a comparable bandwidth for the local quadratic relative to the local linear regression. Note that under the three null hypotheses, $f_{Y|X}\left(\tau|x_0^+\right) = f_{Y|X}\left(\tau|x_0^+\right) = f_{Y|X}\left(\tau|x_0\right)$.

Step 1. We verify that $G_*^R(\tau)$ converges weakly to a continuous Gaussian process over $\mathcal{T}$ under both bandwidth sequences.

Under bandwidth sequence (i), $(\sqrt{nh_{n,\tau}^5}/\sqrt{nb_{n,\tau}^5})\left(D_2^+(\tau) - D_2^-(\tau)\right)$ converges weakly to 0 over $\tau \in \mathcal{T}$. Therefore,

$$G_*^R(\tau) = \hat{f}_{Y|X}(\tau|x_0)\{D_1^+(\tau) - D_1^-(\tau)\} + o_p(1) \Rightarrow G_1(\tau) \text{ over } \tau \in \mathcal{T},$$

where $G_1(\tau)$ is the Gaussian process defined in (S.7).

Under bandwidth sequence (ii), the limit of $\hat{f}_{Y|X}(\tau|x_0)(D_1^+(\tau) - D_1^-(\tau))$ is still given by $G_1(\tau)$. The limit of $\hat{f}_{Y|X}(\tau|x_0)(\sqrt{nh_{n,\tau}^5}/\sqrt{nb_{n,\tau}^5})(D_2^+(\tau) - D_2^-(\tau))$, denoted by $G_2(\tau)$, is a zero-mean Gaussian process with covariance function

$$E\left[G_2(t)G_2(s)\right] = \frac{(t \wedge s - ts)\Gamma^2}{f_X(x_0)\left(\kappa(t)\kappa(s)\right)^{1/2}\left(r(t)r(s)\right)^{5/2}} \int_{-\infty}^{\infty} H_2(t)H_2(s)K\left(\frac{u}{\kappa(t)}\right)K\left(\frac{u}{\kappa(s)}\right)du,$$

where

$$g(\tau)' = \left[1 \;\; \frac{u}{\kappa(\tau)} \;\; \frac{u^2}{\kappa(\tau)^2}\right], r(\tau) = b_{n,\tau}/h_{n,\tau}, \kappa(\tau) = b_{n,\tau}/b_{n,1/2},$$

and

$$H_2(\tau) = \iota_3'\left\{\frac{f_{Y|X}(\tau|x_0)}{f_{Y|X}\left(\tau|x_0^+\right)}(\bar{N}^+)^{-1}I(u \geq 0) - \frac{f_{Y|X}(\tau|x_0)}{f_{Y|X}\left(\tau|x_0^-\right)}(\bar{N}^-)^{-1}\left(1 - I(u \geq 0)\right)\right\}g(\tau).$$

Therefore,

$$G_*^R(\tau) \Rightarrow G_1(\tau) - G_2(\tau) \text{ over } \tau \in \mathcal{T}.$$

Step 2. Denote the simulated version of $G_*^R(\tau)$ by $\hat{S}_*^R(\tau)$. We prove that, if some convergences hold, then $\hat{S}_*^R(\tau)$ converges weakly to the same Gaussian process as given in Step 1, conditionally on the original observations.

We first establish some general results, and then apply them to the two bandwidth sequences (i) and (ii). It is useful to write out the expression of $\hat{S}_*^R(\tau)$ explicitly:

$$\hat{S}_*^R(\tau) = [\widehat{S}_1^+(\tau) - \widehat{S}_1^-(\tau)] - [\widehat{S}_2^+(\tau) - \widehat{S}_2^-(\tau)],$$

where

$$\begin{aligned}
\widehat{S}_1^+(\tau) &= \frac{\hat{f}_{Y|X}(\tau|x_0) f_X(x_0)}{\hat{f}_{Y|X}\left(\tau|x_0^+\right) \hat{f}_X(x_0)} S_1^+(\tau), \qquad \text{(S.17)}\\
\widehat{S}_2^+(\tau) &= r(\tau)^{5/2} \frac{\sqrt{nh_{n,\tau}^5}}{\sqrt{nb_{n,\tau}^5}} \frac{\hat{f}_{Y|X}(\tau|x_0) f_X(x_0)}{\hat{f}_{Y|X}\left(\tau|x_0^+\right) \hat{f}_X(x_0)} S_2^+(\tau),\\
S_1^+(\tau) &= \frac{(nh_{n,\tau})^{-1/2} \sum_{i=1}^n (\tau - 1(u_i - \tau \leq 0)) \Xi_{i,\tau}^+ d_i K_{i,\tau}}{f_X(x_0)},\\
S_2^+(\tau) &= r(\tau)^{-5/2} \Gamma \frac{\iota_3' (\bar{N}^+)^{-1} (nb_{n,\tau})^{-1/2} \sum_{i=1}^n \{\tau - 1(u_i - \tau \leq 0)\} d_i \bar{z}_{i,\tau} \bar{K}_{i,\tau}}{f_X(x_0)},
\end{aligned}$$

and $\widehat{S}_1^-(\tau)$ and $\widehat{S}_2^-(\tau)$ are defined in the same way as $\widehat{S}_1^+(\tau)$ and $\widehat{S}_2^+(\tau)$, except using observations on the left side of the cut-off. The parameter $r(\tau)$ is defined only under bandwidth sequence (ii). It can be set to any finite positive value under bandwidth sequence (i).

For now, assume the following three convergences hold for the sample sequence $(x_1, y_1), (x_2, y_2), ...$:

(C1) $\hat{f}_{Y|X}(\tau|x_0^+) \to f_{Y|X}(\tau|x_0^+)$, $\hat{f}_{Y|X}(\tau|x_0^-) \to f_{Y|X}(\tau|x_0^-)$, and $\hat{f}_{Y|X}(\tau|x_0) \to f_{Y|X}(\tau|x_0)$ uniformly over $\tau \in \mathcal{T}$. In addition, $\hat{f}_X(x_0) \to f_X(x_0)$.

(C2) For any $t, s \in \mathcal{T}$,

$$\frac{t \wedge s - ts}{n (h_{n,t} h_{n,s})^{1/2}} \sum_{i=1}^n \frac{(\Xi_{i,t}^+ d_i K_{i,t} - \Xi_{i,t}^- (1-d_i) K_{i,t})(\Xi_{i,s}^+ d_i K_{i,s} - \Xi_{i,s}^- (1-d_i) K_{i,s})}{f_X(x_0)^2} \to E[G_1(t) G_1(s)]$$

(C3) Under bandwidth sequence (ii), for any $t, s \in \mathcal{T}$,

$$\begin{aligned}
\frac{t \wedge s - ts}{n (b_{n,t} b_{n,s})^{1/2}} \sum_{i=1}^n \Bigg\{ & \frac{\Gamma^2}{r(t)^{5/2} r(s)^{5/2} f_X(x_0)^2} \iota_3' \left[(\bar{N}^+)^{-1} d_i - (\bar{N}^-)^{-1} (1-d_i)\right] \times \\
& \bar{z}_{i,t} \bar{K}_{i,t} \bar{K}_{i,s} \bar{z}_{i,s}' \left[(\bar{N}^+)^{-1} d_i - (\bar{N}^-)^{-1} (1-d_i)\right]' \iota_3 \Bigg\} \to E[G_2(t) G_2(s)].
\end{aligned}$$

We claim that, for every sequence $(x_1, y_1), (x_2, y_2), ...$ that satisfies (C1)-(C3), the following two results always hold:

(R1) The process $S_1^+(\tau) - S_1^-(\tau)$, conditional on $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, converges weakly to $G_1(\tau)$ over $\mathcal{T}$.

(R2) Under bandwidth sequence (ii), the process $S_2^+(\tau) - S_2^-(\tau)$, conditional on $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, converges weakly to $G_2(\tau)$ over $\mathcal{T}$.

Claim (R1) can be proved as below. First, the finite dimensional convergence of $S_1^+(\tau) - S_1^-(\tau)$ follows by applying the Cramer-Wold device conditionally, and then applying (C2). Note that the left-hand side of (C2) equals the covariance of $S_1^+(t)$ and $S_1^+(s)$, conditional on $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$. Second, the stochastic equicontinuity of $S_1^+(\tau) - S_1^-(\tau)$ can be verified by applying the arguments in Lemma B.3 in Qu and Yoon (2015) conditionally, and then applying (C2). Claim (R1) then follows by combining these two results. Claim (R2) can be proved in the same way, using the condition (C3) instead of (C2).

Now, we apply (C1)−(C3) and (R1)−(R2) to the two bandwidth sequences. Under bandwidth sequence (i), (C1) and (R2) imply $\widehat{S}_2^+(\tau) - \widehat{S}_2^-(\tau)$ converges weakly to 0 conditionally. Further, (C1) and (R1) imply $\widehat{S}_1^+(\tau) - \widehat{S}_1^-(\tau)$ converges weakly to $G_1(\tau)$ conditionally. Therefore, $\hat{G}_*^R(\tau)$ converges weakly to $G_1(\tau)$, conditional on $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$. Under bandwidth sequence (ii), (C1) and (R2) imply $\widehat{S}_2^+(\tau) - \widehat{S}_2^-(\tau)$ converges weakly to $G_2(\tau)$ conditionally. Further, (C1) and (R1) imply $\widehat{S}_1^+(\tau) - \widehat{S}_1^-(\tau)$ converges weakly to $G_1(\tau)$ conditionally. Therefore, $\hat{G}_*^R(\tau)$ converges weakly to $G_1(\tau) - G_2(\tau)$, conditional on $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$.

Step 3. We show that (C1)−(C3) hold in probability for the original sample sequence $(x_1, y_1), (x_2, y_2), ...$

For (C1), $\hat{f}_{Y|X}(\tau|x_0^+) \xrightarrow{p} f_{Y|X}(\tau|x_0^+)$ uniformly over $\mathcal{T}$, because $\sqrt{nh_{n,\tau}}(\hat{Q}(\tau|x_0^+) - Q(\tau|x_0^+)) = O_p(1)$ uniformly over $\mathcal{T}$, see (11). Similarly, $\hat{f}_{Y|X}(\tau|x_0^-) \xrightarrow{p} f_{Y|X}(\tau|x_0^-)$ uniformly over $\mathcal{T}$. By the continuous mapping theorem, $\hat{f}_{Y|X}(\tau|x_0)$ converges in probability to $f_{Y|X}(\tau|x_0)$ uniformly over $\mathcal{T}$. Finally, $\hat{f}_X(x_0) \to^p f_X(x_0)$ because $\hat{f}_X(x_0)$ is a standard kernel density estimator.

To prove (C2), it suffices to verify that the expectation of the left-hand side of (C2) converges to $E[G_1(t)G_1(s)]$, and that its variance converges to 0. Because the summands are i.i.d., the expectation of the left hand side is equal to

$$\frac{t \wedge s - ts}{(h_{n,t}h_{n,s})^{1/2}} E\left\{ \frac{(\Xi_{i,t}^+ d_i K_{i,t} - \Xi_{i,t}^-(1-d_i)K_{i,t})(\Xi_{i,s}^+ d_i K_{i,s} - \Xi_{i,s}^-(1-d_i)K_{i,s})}{f_X(x_0)^2} \right\}. \tag{S.18}$$

Consider the following component of (S.18):

$$\begin{aligned} & \frac{t \wedge s - ts}{(h_{n,t}h_{n,s})^{1/2}} E\left\{ \frac{\Xi_{i,t}^+ d_i K_{i,t} \Xi_{i,s}^-(1-d_i)K_{i,s}}{f_X(x_0)^2} \right\} \\ = & \frac{t \wedge s - ts}{(h_{n,t}h_{n,s})^{1/2} f_X(x_0)^2 \left(\mu_0^+\mu_2^+ - (\mu_1^+)^2\right)^2} \int_{-\infty}^{\infty} \left(\mu_2^+ - \left(\frac{x-x_0}{h_{n,t}}\right)\mu_1^+\right) I(x \geq x_0) \\ & \times K\left(\frac{x-x_0}{h_{n,t}}\right)\left(\mu_2^- - \left(\frac{x-x_0}{h_{n,s}}\right)\mu_1^-\right)(1 - I(x \geq x_0)) K\left(\frac{x-x_0}{h_{n,s}}\right) f_X(x)dx. \end{aligned}$$

Let $u = (x - x_0)/h_{n,0.5}$ and apply the mean value theorem. Then, the preceding display converges

to

$$\frac{t \wedge s - ts}{f_X(x_0)(\mu_0^+\mu_2^+ - (\mu_1^+)^2)^2 (\kappa(t)\kappa(s))^{1/2}} \int_{-\infty}^{\infty} \left(\mu_2^+ - \left(\frac{u}{\kappa(t)}\right)\mu_1^+\right) I(u \geq 0) \tag{S.19}$$
$$\times K\left(\frac{u}{\kappa(t)}\right)\left(\mu_2^- - \left(\frac{u}{\kappa(s)}\right)\mu_1^-\right)(1 - I(u \geq 0))\, K\left(\frac{u}{\kappa(s)}\right) du.$$

The remaining components of (S.18) can be analyzed in the same way. Combining these results, it follows that (S.18) converges to $E[G_1(t)G_1(s)]$. To show the variance of the left hand side of (C2) converges to zero, it is sufficient to prove that

$$\frac{(t \wedge s - ts)^2}{n^2 (h_{n,t}h_{n,s})} \sum_{i=1}^{n}\sum_{j=1}^{n} E\left(\frac{(\Xi_{i,t}^+ d_i K_{i,t} - \Xi_{i,t}^-(1-d_i)K_{i,t})(\Xi_{i,s}^+ d_i K_{i,s} - \Xi_{i,s}^-(1-d_i)K_{i,s})}{f_X(x_0)^2}\right)$$
$$\times \left(\frac{(\Xi_{j,t}^+ d_j K_{j,t} - \Xi_{j,t}^-(1-d_j)K_{j,t})(\Xi_{j,s}^+ d_j K_{j,s} - \Xi_{j,s}^-(1-d_j)K_{j,s})}{f_X(x_0)^2}\right) \to (E[G_1(t)G_1(s)])^2,$$

which holds because of the arguments (S.18)-(S.19), the independence of the summands when $i \neq j$, and $nh_{n,0.5} \to \infty$. The convergence in (C3) holds for the same reason as that in (C2); the detail is omitted.

Step 4. We apply a subsequence argument to show that the simulation procedure is weakly consistent.

First, from Step 3, any subsequence of $(x_1, y_1), (x_2, y_2), ...$ contains a further subsequence, such that (C1)−(C3) holds with probability 1, by Theorem 20.5 in Billingsley (1986). Second, from Step 2, conditional on any of such further subsequence, the simulated process $\hat{S}_*^R(\tau)$ converges weakly to the same limit as $G_*^R(\tau)$ does. Therefore, $\hat{S}_*^R(\tau)$, conditional on the original sequence $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, converges weakly to the desired limit in probability. This implies that the simulation procedure is weakly consistent.

S.6 Additional simulation results

This subsection considers three issues. First, it compares the power of the score and Wald tests under an ideal simulation setting. Second, it evaluates the effect of estimating the conditional density on the size of the Wald tests. Third, it reports the rejection rates at the 5% nominal level for Models 1−4 considered in Section 8.

Power comparison. Tests may display different power if they use distinct bandwidths or different bias correction methods. To exclude such effects, we compare the score test and the Wald test for the treatment significance hypothesis using the same bandwidth without bias estimation. The sample size is 1000 and the bandwidth at the median is 0.4. The data generating processes are

Models 1 and 2. The rejection frequencies at the 10% level are reported in Table A1. The values are comparable between the two tests for all the cases considered. This confirms the results from the local power analysis.

Conditional density estimation. Section 8 documents that the Wald tests have less stable size properties compared to those of the score test, especially when the sample size is small. Here, we examine the extent to which this is because the Wald tests require estimating the conditional density. We repeat the same procedures as in Section 8, but using the true conditional densities instead of the estimated densities. The data generating processes are Models 1 and 2 for which all three tests are valid. The sample size is $n = 500$ and the nominal level is 10%. Table A2 shows the results. Compared with Tables 2−4, the values are now consistently close to the nominal level. They are also comparable to those of the score test. Therefore, estimating conditional densities accounts for most of the size distortions in small samples.

Empirical sizes at the 5% nominal level. Section 8 only reports sizes at the 10% level. To complement these results, here we also report the sizes the 5% level. The results are shown in Tables A3−A6. Overall, the same patterns as in Tables 2−4 and 9 are observed. The conclusions therefore remain the same.

Table A1: Power of Score and Wald tests using the same bandwidth (10%).

Test	**Model 1**				**Model 2**			
	c_h=0.3	0.6	1.0	2.0	c_h=0.3	0.6	1.0	2.0
Score	0.265	0.663	0.948	1.000	0.386	0.741	0.975	1.000
Wald	0.293	0.648	0.922	1.000	0.336	0.693	0.941	1.000

Empirical rejection frequencies based on 2000 repetitions. The sample size $n = 1000$ and the bandwidth at the median is fixed at 0.4.

Table A2: The Size of Wald tests using true conditional density functions (10%).

Methods	**Model 1**			**Model 2**		
	TS	**TH**	**TU**	**TS**	**TH**	**TU**
Wald						
$h_{0.5}^{cv}$	0.102	0.108	0.104	0.101	0.101	0.076
$h_{0.5}^{bdy}$	0.105	0.100	0.111	0.110	0.096	0.090
$h_{0.5}^{ik}$	0.105	0.106	0.106	0.102	0.098	0.102
Wald Robust						
$h_{0.5}^{cv}$	0.094	0.094	0.098	0.094	0.089	0.097
$h_{0.5}^{bdy}$	0.090	0.090	0.092	0.094	0.076	0.100
$h_{0.5}^{ik}$	0.096	0.081	0.098	0.100	0.074	0.096
Wald Robust EC						
$h_{0.5}^{cv}$	0.106	0.100	0.098	0.100	0.090	0.103
$h_{0.5}^{bdy}$	0.103	0.096	0.106	0.108	0.082	0.104
$h_{0.5}^{ik}$	0.100	0.096	0.102	0.100	0.090	0.102

Empirical rejection frequencies based on 2000 repetitions. The sample size is $n = 500$. **TS**, **TH**, and **TU** stand for the treatment significance/homogeneity/unambiguity hypotheses.

Table A3: The Size of Tests for the Treatment Significance Hypothesis (Models 1 & 2).

Tests	**Model 1**			**Model 2**		
	n=500	n=1000	n=2000	n=500	n=1000	n=2000
Score						
$h_{0.5}^{cvi}$	0.054	0.042	0.048	0.057	0.050	0.048
$h_{0.5}^{int}$	0.054	0.058	0.056	0.062	0.062	0.058
Wald						
$h_{0.5}^{cv}$	0.090	0.067	0.056	0.098	0.094	0.072
$h_{0.5}^{bdy}$	0.101	0.074	0.058	0.116	0.084	0.062
$h_{0.5}^{ik}$	0.130	0.084	0.059	0.132	0.092	0.061
Wald Robust						
$h_{0.5}^{cv}$	0.081	0.051	0.045	0.084	0.060	0.041
$h_{0.5}^{bdy}$	0.085	0.059	0.042	0.091	0.062	0.042
$h_{0.5}^{ik}$	0.102	0.074	0.042	0.109	0.074	0.040
Wald Robust EC						
$h_{0.5}^{cv}$	0.088	0.060	0.054	0.097	0.068	0.050
$h_{0.5}^{bdy}$	0.094	0.064	0.052	0.112	0.070	0.051
$h_{0.5}^{ik}$	0.113	0.074	0.049	0.122	0.078	0.048

Note. The table reports rejection frequencies at the **5 percent** nominal level based on 2000 replications. "Wald", "Wald Robust" and "Wald Robust EC" denote tests constructed assuming a continuous second order derivative at the cutoff, allowing a discontinuous second order derivative whose magnitude of discontinuity can vary freely across the quantiles, and allowing a discontinuous second order derivative whose magnitude of discontinuity remains constant across the quantiles. See the footnote of Table 1 in the main text for the definitions of the bandwidth parameters.

Table A4: The Size of Tests for the Treatment Unambiguity Hypothesis (Models 1 & 2).

Tests	**Model 1**			**Model 2**		
	n=500	n=1000	n=2000	n=500	n=1000	n=2000
Wald						
$h_{0.5}^{cv}$	0.086	0.052	0.046	0.062	0.044	0.031
$h_{0.5}^{bdy}$	0.088	0.057	0.047	0.078	0.052	0.040
$h_{0.5}^{ik}$	0.100	0.064	0.055	0.102	0.062	0.047
Wald Robust						
$h_{0.5}^{cv}$	0.070	0.049	0.040	0.064	0.046	0.037
$h_{0.5}^{bdy}$	0.077	0.053	0.046	0.088	0.053	0.044
$h_{0.5}^{ik}$	0.082	0.053	0.047	0.087	0.058	0.042
Wald Robust EC						
$h_{0.5}^{cv}$	0.078	0.054	0.052	0.078	0.052	0.044
$h_{0.5}^{bdy}$	0.082	0.056	0.048	0.092	0.058	0.042
$h_{0.5}^{ik}$	0.088	0.058	0.046	0.091	0.057	0.044

Note. The nominal level is **5 percent**. See Table A3.

Table A5: The Size of Tests for the Treatment Homogeneity Hypothesis (Models 1 & 2).

Tests	**Model 1**			**Model 2**		
	n=500	n=1000	n=2000	n=500	n=1000	n=2000
Wald						
$h_{0.5}^{cv}$	0.084	0.060	0.046	0.088	0.064	0.052
$h_{0.5}^{bdy}$	0.082	0.062	0.048	0.094	0.071	0.050
$h_{0.5}^{ik}$	0.102	0.073	0.052	0.110	0.068	0.057
Wald Robust						
$h_{0.5}^{cv}$	0.077	0.049	0.042	0.077	0.048	0.038
$h_{0.5}^{bdy}$	0.078	0.053	0.038	0.080	0.052	0.040
$h_{0.5}^{ik}$	0.096	0.066	0.043	0.102	0.070	0.046
Wald Robust EC						
$h_{0.5}^{cv}$	0.072	0.053	0.042	0.074	0.056	0.044
$h_{0.5}^{bdy}$	0.064	0.054	0.041	0.076	0.062	0.042
$h_{0.5}^{ik}$	0.088	0.064	0.046	0.098	0.059	0.045

Note. The nominal level is **5 percent**. See Table A3.

Table A6: The Size of Robust Tests in Models 3 & 4.

Tests	**Model 3**			**Model 4**		
	n=500	n=1000	n=2000	n=500	n=1000	n=2000
Treatment Significance:						
Wald Robust						
$h_{0.5}^{cv}$	0.110	0.094	0.086	0.125	0.075	0.054
$h_{0.5}^{bdy}$	0.106	0.063	0.042	0.124	0.060	0.046
$h_{0.5}^{ik}$	0.110	0.062	0.055	0.105	0.058	0.036
Wald Robust EC						
$h_{0.5}^{cv}$	0.123	0.100	0.100	0.133	0.078	0.060
$h_{0.5}^{bdy}$	0.122	0.072	0.051	0.136	0.079	0.060
$h_{0.5}^{ik}$	0.121	0.084	0.064	0.120	0.064	0.041
Treatment Unambiguity:						
Wald Robust						
$h_{0.5}^{cv}$	0.029	0.026	0.020	0.078	0.052	0.052
$h_{0.5}^{bdy}$	0.048	0.038	0.028	0.076	0.050	0.048
$h_{0.5}^{ik}$	0.052	0.036	0.018	0.066	0.036	0.031
Wald Robust EC						
$h_{0.5}^{cv}$	0.032	0.027	0.019	0.082	0.061	0.057
$h_{0.5}^{bdy}$	0.047	0.046	0.034	0.083	0.056	0.048
$h_{0.5}^{ik}$	0.057	0.042	0.022	0.067	0.038	0.032
Treatment Homogeneity:						
Wald Robust						
$h_{0.5}^{cv}$	0.072	0.044	0.050	0.104	0.068	0.061
$h_{0.5}^{bdy}$	0.080	0.052	0.054	0.096	0.066	0.054
$h_{0.5}^{ik}$	0.080	0.048	0.050	0.077	0.066	0.044
Wald Robust EC						
$h_{0.5}^{cv}$	0.068	0.052	0.048	0.098	0.082	0.064
$h_{0.5}^{bdy}$	0.088	0.052	0.060	0.100	0.074	0.058
$h_{0.5}^{ik}$	0.086	0.056	0.045	0.091	0.066	0.050

Note. The nominal level is **5 percent**. See Table A3.