Inference in Regression Discontinuity Designs with a Discrete Running Variable: Supplemental Materials

Michal Kolesár* Christoph Rothe†

March 27, 2018

This supplement is organized as follows. Section S1 generalizes the setup in Section III and gives proofs of results given in that section. Section S2 gives proofs of results in Section IV. Section S3 shows the fit of the specifications from the CPS placebo study considered in Section I, and Section S4 gives additional details and results for the placebo study, including results on the performance of the honest CIs. Section S5 considers an additional Monte Carlo study. Finally, Section S6 constructs an estimate for a lower bound on the smoothness constant K from Section IV.A.

S1. PROOFS OF RESULTS IN SECTION III

The claims in Section III follow directly from general results on the properties of $\widehat{\sigma}^2_{\text{CRV}}$ that are given in the following subsection. The proofs of these results are given in turn in Sections A.2–A.4. To state these results, we use the notation $\operatorname{diag}\{a_g\}$ to denote a diagonal matrix with diagonal elements given by $a_1, \dots, a_G$, and $\operatorname{vec}\{a_g\} = (a_1', \dots, a_G')'$.

S1.1. Properties of $\widehat{\sigma}^2_{\text{CRV}}$ under General Conditions

In this subsection, we consider a setup that is slightly more general than that in Section 3, in that it also allows the bandwidth h to change with the sample size. For convenience, the following assumption summarizes this more general setup.

Assumption 1 (Model)**.** *For each N, the data $\{Y_i, X_i\}_{i=1}^N$ are i.i.d., distributed according to a law P_N. Under P_N, the marginal distribution of X_i is discrete with $G = G_- + G_+$ support points denoted $x_1 < \dots < x_{G^-} < 0 \leq x_{G^-+1} < \dots < x_G$. Let $\mu(x) = \mathbb{E}_N(Y_i \mid X_i = x)$ denote the CEF under P_N. Let $\varepsilon_i = Y_i - \mu(X_i)$, and let $\sigma_g^2 = \mathbb{V}_N(\varepsilon_i \mid X_i = x_g)$ denote its conditional variance. Let $h = h_N$ denote a non-random bandwidth sequence, and let $\mathcal{G}_h \subseteq \{1, \dots, G\}$ denote the indices for which $|x_g| \leq h$, with G_h^+ and G_h^- denoting the number of elements in $\mathcal{G}_h$ above and below zero. Let $\pi_g = P_N(X_i = x_g)$, $\pi = P_N(|X_i| \leq h)$, and $N_h = \sum_{i=1}^N \mathbb{I}\{|X_i| \leq h\}$. For a fixed integer $p \geq 0$, define*

$$m(x) = (\mathbb{I}\{x \geq 0\}, 1, x, \dots, x^p, \mathbb{I}\{x \geq 0\} x, \dots, \mathbb{I}\{x \geq 0\} x^p)',$$

*Woodrow Wilson School and Department of Economics, Princeton University, mkolesar@princeton.edu
†Department of Economics, University of Mannheim, rothe@vwl.uni-mannheim.de

$M_i = \mathbb{I}\{|X_i| \leq h\} m(X_i)$*, and* $m_g = \mathbb{I}\{|x_g| \leq h\} m(x_g)$*. Let* $\widehat{Q} = \frac{1}{N_h}\sum_{i=1}^{n} M_i M_i'$*, and* $Q_N = \mathbb{E}_N(M_i M_i')/\pi$*. Let* $\theta_h = Q_N^{-1}\mathbb{E}_N(M_i Y_i)/\pi$*, and denote its first element my* τ_h*. Let* $\widehat{\theta} = \widehat{Q}^{-1}\frac{1}{N_h}\sum_{i=1}^{N} M_i Y_i$*, and denote its first element by* $\widehat{\tau}$*. Define* $\delta(x) = \mu(x) - m(x)'\theta_h$*, and* $u_i = Y_i - m(X_i)'\theta_h = \delta(X_i) + \epsilon_i$*. Define* $\Omega = \mathbb{E}_N(u_i^2 M_i M_i')/\pi = \sum_{g=1}^{G}(\sigma_g^2 + \delta^2(x_g))Q_g$*, where* $Q_g = \frac{\pi_g}{\pi} m_g m_g'$*.*

Note that the setup allows various quantities that depend on P_N and h to change with N, such as the number of support points G, their locations x_g, the conditional expectation function $\mu(x)$, or the specification errors $\delta(X_i)$.

Assumption 2 (Regularity conditions)**.** *(i)* $\sup_N \max_{g\in\{1,\ldots,G\}} \mathbb{E}_N(\epsilon_i^4 \mid X_i = x_g) < \infty$*,* $\det(H^{-1}Q_N H^{-1}) = \det(\sum_{g\in\mathcal{G}_h} \frac{\pi_g}{\pi} m(x_g/h)m(x_g/h)') > C$ *for some* $C > 0$ *that does not depend on* N*, where* $H = \operatorname{diag}\{m(h)\}$*,* $N\pi \to \infty$*, and the limit* $\lim_{N\to\infty} H^{-1}Q_N H^{-1}$ *exists. (ii)* $\sup_N \max_{g\in\mathcal{G}_h} \delta(x_g) < \infty$*; and the limit* $\lim_{N\to\infty} H^{-1}\Omega H^{-1}$ *exists.*

The assumption ensures that that bandwidth shrinks to zero slowly enough so that the number of effective observations $N\pi$ increases to infinity, and that the number of effective support points $G_h = G_h^+ + G_h^-$ is large enough so that the parameter θ_h and the asymptotic variance of $\widehat{\theta}$ remain well-defined with well-defined limits. We normalize Q_N and Ω by the inverse of H since if $h \to 0$, their elements converge at different rates.

Our first result is an asymptotic approximation in which G_h^+ and G_h^- are fixed as the sample size increases. Let $B_1, \ldots, B_G$ be a collection of random vectors such that $\operatorname{vec}\{B_g\} \sim \mathcal{N}(0, V)$, with

$$V = \frac{1}{\pi}\operatorname{diag}\{\pi_g(\sigma_g^2 + \delta(x_g)^2)\} - \frac{1}{\pi}\operatorname{vec}\{\pi_g\delta(x_g)\}\operatorname{vec}\{\pi_g\delta(x_g)\}'.$$

Note that if $|x_g| > h$, then $B_g = 0$ and $Q_g = 0$, and that the limiting distribution of the statistic $\sqrt{N_h}(\widehat{\tau} - \tau_h)$ coincides with the distribution of $e_1'Q_N^{-1}\sum_{g=1}^{G} m_g B_g$. Finally, define

$$W_g = e_1'Q_N^{-1}m_g\left(B_g - \frac{\pi_g}{\pi}m_g'Q_N^{-1}\sum_{j=1}^{G} m_j B_j + (N/\pi)^{1/2}\pi_g\delta(x_g)\right).$$

With this notation, we obtain the following generic result.

Theorem S1. *Suppose that Assumptions 1 and 2 hold. Suppose also that, as* $N \to \infty$*, (i)* G_h^+ *and* G_h^- *are fixed; and (ii) the limit of* V *exists. Then*

$$\widehat{\sigma}^2_{\text{CRV}} \stackrel{d}{=} (1 + o_{P_N}(1))\sum_{g=1}^{G} W_g^2.$$

Our second result is an asymptotic approximation in which the number of support points of the running variable (or, equivalently, the number of “clusters”) that are less than h away from the threshold increases with the sample size.

Theorem S2. *Suppose that Assumptions 1 and 2 hold. Suppose also that, as $N \to \infty$, $G_h \to \infty$ and $\max_{g \in \mathcal{G}_h} \pi_g/\pi \to 0$. Then*

$$\widehat{\sigma}^2_{\text{CRV}} = (1 + o_{P_N}(1)) e_1' Q_N^{-1} \left(\Omega + (N-1) \sum_{g=1}^{G} Q_g \cdot \pi_g \delta(x_g)^2. \right) Q_N^{-1} e_1.$$

The assumption that $\max_{g \in \mathcal{G}_h} \pi_g/\pi \to 0$ ensures that each "cluster" comprises a vanishing fraction of the effective sample size.

S1.2. **Auxiliary Lemma**

Here we state an intermediate result that is used in the proofs of Theorem 1 and 2 below, and that shows that $\widehat{\sigma}^2_{\text{EHW}}$ is consistent for the asymptotic variance of $\widehat{\theta}$.

Lemma S1. *Suppose that Assumptions 1 and 2 (i) hold. Then*

$$\frac{N_h/N}{\pi} = 1 + o_{P_N}(1), \tag{S1}$$

$$H^{-1}\widehat{Q}H^{-1} - H^{-1}Q_N H^{-1} = o_{P_N}(1). \tag{S2}$$

If, in addition, Assumption 2 (ii) holds, then

$$\sqrt{N_h} H(\widehat{\theta} - \theta_h) \overset{d}{=} H Q_N^{-1} S + o_{P_N}(1), \tag{S3}$$

where $S \sim \mathcal{N}(0, \Omega)$.

Let $n_g = \sum_{i=1}^{N} \mathbb{I}\{X_i = x_g\}$, $\widehat{q}_g = H\widehat{Q}^{-1}Hm(x_g/h)\mathbb{I}\{|x_g| \leq h\}$, and let

$$A_g = \frac{\mathbb{I}\{|x_g| \leq h\}}{\sqrt{N\pi}} \sum_{i=1}^{N} \left(\mathbb{I}\{X_i = x_g\} \varepsilon_i + (\mathbb{I}\{X_i = x_g\} - \pi_g)\delta(x_g) \right).$$

Then $H^{-1}\sum_{g=1}^{G} m_g A_g \overset{d}{=} H^{-1}S + o_{P_N}(1)$, and

$$\widehat{\sigma}^2_{\text{CRV}} = (1 + o_{P_N}(1)) \sum_{g=1}^{G} (e_1'\widehat{q}_g)^2 \left(A_g - \frac{n_g}{N_h} \widehat{q}_g' \sum_{j=1}^{G} m(x_j/h) A_j + \frac{\sqrt{N}\pi_g \delta(x_g)}{\sqrt{\pi}} \right)^2. \tag{S4}$$

Furthermore, $\widehat{\sigma}^2_{\text{EHW}} = e_1' Q_N^{-1} \Omega Q_N^{-1} e_1 + o_{P_N}(1)$.

Proof. We have $\mathbb{V}_N(N_h/N) = \pi(1-\pi)/N \leq \pi/N$. Therefore, by Markov's inequality, $N\pi \to \infty$ implies $\frac{N_h/N}{\pi} = \mathbb{E}_N(N_h/(N\pi)) + o_{P_N}(1) = 1 + o_{P_N}(1)$, which proves (S1). Secondly, since elements of $H^{-1}M_i$ are bounded by $\mathbb{I}\{|X_i| \leq h\}$, the second moment of any element of $\frac{N_h}{N\pi}H^{-1}\widehat{Q}H^{-1} - H^{-1}Q_N H^{-1} = \frac{1}{N\pi}\sum_{i=1}^{n} H^{-1}(M_i M_i' - E[M_i M_i'])H^{-1}$ is bounded by $1/(N\pi)$, which converges to zero by assumption. Thus, by Markov's inequality, $\frac{N_h}{N\pi}H^{-1}\widehat{Q}H^{-1} - H^{-1}Q_N H^{-1} = o_{P_N}(1)$. Combining this result with (S1) and the fact that $H^{-1}Q_N H^{-1}$ is bounded then yields (S2).

Next note that since $\sum_{g=1}^{N} \pi_g m_g \delta(x_g) = 0$, $H^{-1}\sum_{g=1}^{G} m_g A_g = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{\pi}}H^{-1}M_i u_i$, and that by the central limit theorem, $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{\pi}}H^{-1}M_i u_i \overset{d}{=} H^{-1}S + o_{P_N}(1)$. Therefore,

$$\sqrt{N_h}H(\widehat{\theta}-\theta_h) = \sqrt{\frac{\pi N}{N_h}}(H^{-1}\widehat{Q}H^{-1})^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{\pi}}H^{-1}M_i u_i \overset{d}{=} HQ_N^{-1}S + o_{P_N}(1),$$

as claimed. Next, we prove (S4). Let $J_g = \sum_{i=1}^{N}\mathbb{I}\{X_i = x_g\}\widehat{u}_i M_i$. Then by (S1), the cluster-robust variance estimator can be written as

$$\widehat{\sigma}^2_{\text{CRV}} = e_1'\widehat{Q}^{-1}\frac{1}{N_h}\sum_{g=1}^{G}J_g J_g'\widehat{Q}^{-1}e_1 = (1+o_{P_N}(1))\sum_{g=1}^{G}\left(\frac{1}{\sqrt{N\pi}}e_1'\widehat{Q}^{-1}J_g\right)^2.$$

The expression in parentheses can be decomposed as

$$\begin{aligned}\frac{1}{\sqrt{N\pi}}e_1'\widehat{Q}^{-1}J_g &= e_1'\widehat{q}_g\left(A_g + N\pi_g\delta(x_g)/\sqrt{N\pi} - n_g m_g'(\widehat{\theta}-\theta_h)/\sqrt{N\pi}\right)\\ &= e_1'\widehat{q}_g\left(A_g + N\pi_g\delta(x_g)/\sqrt{N\pi} - (n_g/N_h)\widehat{q}_g'H^{-1}\sum_{j=1}^{G}m_j A_j\right),\end{aligned}$$

which yields the result.

It remains to prove consistency of $\widehat{\sigma}^2_{\text{EHW}}$. To this end, using (S1), decompose

$$H^{-1}\widehat{\Omega}_{\text{EHW}}H^{-1} = (1+o_{P_N}(1))\frac{1}{N\pi}\sum_{i=1}^{N}\widehat{u}_i^2 H^{-1}M_i M_i' H^{-1} = (1+o_{P_N}(1))(\mathcal{C}_1+\mathcal{C}_2+\mathcal{C}_3),$$

where $\mathcal{C}_1 = \frac{1}{N\pi}\sum_{i=1}^{N}u_i^2 H^{-1}M_i M_i' H^{-1}$, $\mathcal{C}_2 = \frac{1}{N\pi}\sum_{i=1}^{N}(M_i'(\widehat{\theta}-\theta_h))^2 H^{-1}M_i M_i' H^{-1}$, and $\mathcal{C}_3 = -\frac{2}{N\pi}\sum_{i=1}^{N}u_i M_i'(\widehat{\theta}-\theta_h)H^{-1}M_i M_i' H^{-1}$. Since elements of M_i are bounded by $\mathbb{I}\{|X_i| \leq h\}$, variance of $\mathcal{C}_1$ is bounded by $\mathbb{E}[u_i^4\mathbb{I}\{|X|_i \leq h\}]/N\pi^2 = o_{P_N}(1)$, so that by Markov's inequality, $\mathcal{C}_1 = E[\mathcal{C}_1] + o_{P_N}(1) = H^{-1}\Omega H^{-1} + o_{P_N}(1)$. Similarly, all elements of $\mathcal{C}_2$ are bounded, $\frac{1}{N\pi}\sum_{i=1}^{N}(M_i'(\widehat{\theta}-\theta_h))^2 = o_{P_N}(1)$ by (S1)–(S3). Finally, $\mathcal{C}_3 = o_{P_N}(1)$ by Cauchy-Schwarz inequality. Thus,

$$H^{-1}\widehat{\Omega}_{\text{EHW}}H^{-1} = H^{-1}\Omega H^{-1} + o_{P_N}(1), \tag{S5}$$

and consistency of $\widehat{\sigma}^2_{\text{EHW}}$ then follows by combining this result with (S2). □

S1.3. **Proof of Theorem S1**

Let $q_g = HQ_N^{-1}Hm(x_g/h)\mathbb{I}\{|x_g| \leq h\}$, and define $\widehat{q}_g$, A_g, and n_g as in the statement of Lemma S1. By Lemma S1, $\widehat{q}_g = q_g(1+o_{P_N}(1))$, and by Markov's inequality and Equation (S1), $n_g/N_h = \pi_g/\pi + o_{P_N}(1)$ for $g \in \mathcal{G}_h$. Combining these results with Equation (S4), it follows that the cluster-robust variance estimator satisfies

$$\widehat{\sigma}^2_{\text{CRV}} = (1+o_{P_N}(1))\sum_{g=1}^{G}(e_1'q_g)^2\left(A_g + \sqrt{N\pi}\frac{\pi_g}{\pi}\delta(x_g) - \frac{\pi_g}{\pi}q_g'\sum_{j=1}^{G}m(x_j/h)A_j\right)^2,$$

To prove the theorem, it therefore suffices to show that

$$e_1' q_g \left(A_g - \frac{\pi_g}{\pi} q_g' \sum_{j=1}^{G} m(x_j/h) A_j + \sqrt{N\pi} \frac{\pi_g}{\pi} \delta(x_g) \right) = W_g(1 + o_{P_N}(1)). \tag{S6}$$

This follows from Slutsky's lemma and the fact that by the central limit theorem,

$$\text{vec}\{A_g\} \overset{d}{=} \text{vec}\{B_g\}(1 + o_{P_N}(1)). \tag{S7}$$

S1.4. **Proof of Theorem S2**

Throughout the proof, write $a \preceq b$ to denote $a < Cb$ for some constant C that does not depend on N. By Equation (S4) in Lemma S1, we can write the cluster-robust estimator as $\widehat{\sigma}^2_{\text{CRV}} = (\mathcal{C}_1 + \mathcal{C}_2 + \mathcal{C}_3)(1 + o_{P_N}(1))$, with

$$\begin{aligned}
\mathcal{C}_1 &= \sum_{g=1}^{G} (e_1' \widehat{q}_g)^2 \left(A_g + (N/\pi)^{1/2} \pi_g \delta(x_g) \right)^2, \\
\mathcal{C}_2 &= \widehat{S}' H^{-1} \sum_{g=1}^{G} (\widehat{q}_g' e_1)^2 \frac{n_g^2}{N_h^2} \widehat{q}_g \widehat{q}_g' \cdot H^{-1} \widehat{S}, \\
\mathcal{C}_3 &= -2 \sum_{g=1}^{G} (e_1' \widehat{q}_g)^2 \left(A_g + (N/\pi)^{1/2} \pi_g \delta(x_g) \right) \frac{n_g}{N_h} \widehat{q}_g' H^{-1} \widehat{S},
\end{aligned}$$

where $\widehat{S} = \sum_{j=1}^{G} m_j A_j$, and n_g, A_g and $\widehat{q}_g$ are defined in the statement of the Lemma.

We first show that $\mathcal{C}_2 = o_{P_N}(1)$. Since $H^{-1}\widehat{S} = O_{P_N}(1)$ by Lemma S1, it suffices to show that

$$\sum_{g=1}^{G} (\widehat{q}_g' e_1)^2 \frac{n_g^2}{N_h^2} \widehat{q}_g \widehat{q}_g' = o_{P_N}(1).$$

To this end, note that since elements of $m(x_g/h)$ are bounded by 1, for any j, by Cauchy-Schwarz inequality, $|\widehat{q}_g' e_j| \leq \|e_j' H \widehat{Q}^{-1} H\|_2 \sqrt{2(p+1)}$, where $\|v\|_2$ denotes the Euclidean norm of a vector v. By Lemma S1, $\|e_j' H \widehat{Q}^{-1} H\|_2 = O_{P_N}(1)$ and $N_h/\pi N = 1 + o_{P_N}(1)$ so that

$$\left| \sum_{g=1}^{G} (\widehat{q}_g' e_1)^2 \frac{n_g^2}{N_h^2} e_j \widehat{q}_g \widehat{q}_g' e_k \right| \leq O_{P_N}(1) \sum_{g \in \mathcal{G}_h} \frac{n_g^2}{N_h^2} = O_{P_N}(1) \sum_{g \in \mathcal{G}_h} \frac{n_g^2}{\pi^2 N^2}.$$

Now, since $\mathbb{E}_N(n_g^2) = N\pi_g(1 - \pi_g) + N^2\pi_g^2$, and $\sum_{g \in \mathcal{G}_h} \pi_g = \pi$,

$$\mathbb{E}_N \sum_{g \in \mathcal{G}_h} \frac{n_g^2}{N^2 \pi^2} = \sum_{g \in \mathcal{G}_h} \frac{\pi_g(1 - \pi_g)}{N\pi^2} + \sum_{g \in \mathcal{G}_h} \frac{\pi_g^2}{\pi^2} \leq \left(\frac{1}{N\pi} + \frac{\max_{g \in \mathcal{G}_h} \pi_g}{\pi} \right) \sum_{g \in \mathcal{G}_h} \frac{\pi_g}{\pi} \to 0.$$

Therefore, by Markov's inequality, $\sum_{g \in \mathcal{G}_h} \frac{n_g^2}{\pi^2 N^2} = o_{P_N}(1)$, so that $\mathcal{C}_2 = o_{P_N}(1)$ as claimed.

Next, consider $\mathcal{C}_1$. Let $q_g = HQ_N^{-1}Hm(x_g/h)\mathbb{I}\{|x_g| \leq h\}$. We have

$$\begin{aligned}
\mathcal{C}_1 &= \frac{1}{N\pi}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{g=1}^{G}(e_1'\widehat{q}_g)^2\mathbb{I}\{X_i = x_g\}\mathbb{I}\{X_j = x_g\}(\varepsilon_i + \delta(x_g))(\varepsilon_j + \delta(x_g)) \\
&= (1 + o_{P_N}(1))\frac{1}{N\pi}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{g=1}^{G}(e_1'q_g)^2\mathbb{I}\{X_i = x_g\}\mathbb{I}\{X_j = x_g\}(\varepsilon_i + \delta(x_g))(\varepsilon_j + \delta(x_g)) \\
&= (1 + o_{P_N}(1))\left(\mathcal{C}_{11} + 2(\mathcal{C}_{12} + \mathcal{C}_{13} + \mathcal{C}_{14} + \mathcal{C}_{15} + \mathcal{C}_{16})\right),
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{C}_{11} &= \frac{1}{N\pi}\sum_{i=1}^{N}\sum_{g=1}^{G}(e_1'q_g)^2\mathbb{I}\{X_i = x_g\}(\varepsilon_i + \delta(x_g))^2, \\
\mathcal{C}_{12} &= \frac{1}{N\pi}\sum_{i=1}^{N}\sum_{j=1}^{i-1}\sum_{g=1}^{G}(e_1'q_g)^2\mathbb{I}\{X_i = x_g\}\mathbb{I}\{X_j = x_g\}\varepsilon_i\varepsilon_j, \\
\mathcal{C}_{13} &= \frac{1}{N\pi}\sum_{i=1}^{N}\sum_{j=1}^{i-1}\sum_{g=1}^{G}(e_1'q_g)^2\mathbb{I}\{X_i = x_g\}\mathbb{I}\{X_j = x_g\}\varepsilon_j\delta(x_g), \\
\mathcal{C}_{14} &= \frac{1}{N\pi}\sum_{i=1}^{N}\sum_{j=1}^{i-1}\sum_{g=1}^{G}(e_1'q_g)^2\mathbb{I}\{X_i = x_g\}\mathbb{I}\{X_j = x_g\}\varepsilon_i\delta(x_g), \\
\mathcal{C}_{15} &= \frac{1}{N\pi}\sum_{i=1}^{N}\sum_{j=1}^{i-1}\sum_{g=1}^{G}(e_1'q_g)^2\mathbb{I}\{X_i = x_g\}(\mathbb{I}\{X_j = x_g\} - \pi_g)\delta(x_g)^2, \\
\mathcal{C}_{16} &= \frac{1}{N\pi}\sum_{g=1}^{G}\sum_{i=1}^{N}(i-1)(e_1'q_g)^2\mathbb{I}\{X_i = x_g\}\pi_g\delta(x_g)^2.
\end{aligned}$$

We have

$$\mathbb{E}_N(\mathcal{C}_{11}) = \frac{1}{\pi}\sum_{g=1}^{G}(e_1'q_g)^2\pi_g(\sigma_g^2 + \delta(x_g)^2) = e_1'Q_N^{-1}\Omega Q_N^{-1}e_1,$$

and

$$\mathbb{V}_N(\mathcal{C}_{11}) \leq \frac{1}{N\pi^2}\sum_{g=1}^{G}(e_1'q_g)^4\pi_g\mathbb{E}_N[(\varepsilon_i + \delta(x_g))^4 \mid X_i = x_g] \preceq \frac{\sum_{g\in\mathcal{G}_h}\pi_g}{N\pi^2} = \frac{1}{N\pi} \to 0.$$

Next, $\mathbb{E}_N(\mathcal{C}_{12}) = 0$, and

$$\mathbb{V}_N(\mathcal{C}_{12}) = \frac{N-1}{2N\pi^2}\sum_{g=1}^{G}(e_1'q_g)^4\pi_g^2\sigma_g^2\sigma_g^2 \preceq \frac{\max_g \pi_g\sum_{g=1}^{G}\pi_g}{\pi^2} = \frac{\max_g \pi_g}{\pi} \to 0.$$

The expectations for the remaining terms satisfy $\mathbb{E}_N(\mathcal{C}_{13}) = \mathbb{E}_N(\mathcal{C}_{14}) = \mathbb{E}_N(\mathcal{C}_{15}) = 0$, and

$$\mathbb{E}_N(\mathcal{C}_{16}) = \frac{N-1}{2\pi} \sum_{g=1}^{G} (e_1' q_g)^2 \pi_g^2 \delta(x_g)^2.$$

The variances of $\mathcal{C}_{13}, \dots, \mathcal{C}_{16}$ are all of smaller order than this expectation:

$$\begin{aligned}
\mathbb{V}_N(\mathcal{C}_{13}) &= \frac{1}{N^2\pi^2} \sum_{g}^{G} \sum_{i,k=1}^{N} \sum_{j=1}^{\min\{i,k\}-1} (e_1' q_g)^4 \pi_g^3 \sigma_g^2 \delta(x_g)^2 \preceq \frac{N \max_g \pi_g}{\pi^2} \sum_{g}^{G} (e_1' q_g)^2 \pi_g^2 \delta(x_g)^2 \\
&= o(\mathbb{E}_N(\mathcal{C}_{16})) \\
\mathbb{V}_N(\mathcal{C}_{14}) &= \frac{1}{N^2\pi^2} \sum_{i=1}^{N} \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} \sum_{g=1}^{G} (e_1' q_g)^4 \pi_g^3 \sigma_g^2 \delta(x_g)^2 = o(\mathbb{E}_N(\mathcal{C}_{16})) \\
\mathbb{V}_N(\mathcal{C}_{15}) &= \frac{1}{N^2\pi^2} \sum_{i=1}^{N} \sum_{k=1}^{N} \sum_{j=1}^{\min\{i,k\}-1} \sum_{g,f=1}^{G} \left(\mathbb{I}\{g=f\}\pi_g - \pi_g \pi_f\right) \pi_g \pi_f (e_1' q_g)^2 (e_1' q_f)^2 \delta(x_f)^2 \delta(x_g)^2 \\
&\leq \frac{1}{N^2\pi^2} \sum_{i=1}^{N} \sum_{k=1}^{N} \sum_{j=1}^{\min\{i,k\}-1} \sum_{g=1}^{G} \pi_g^3 (e_1' q_g)^4 \delta(x_g)^4 = o(\mathbb{E}_N(\mathcal{C}_{16})),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{V}_N(\mathcal{C}_{16}) &= \frac{1}{N^2\pi^2} \sum_{g=1}^{G} \sum_{f=1}^{G} \sum_{i=1}^{N} (i-1)^2 (\mathbb{I}\{g=f\}\pi_g - \pi_g\pi_f)\pi_g\pi_f \delta(x_g)^2 \delta(x_f)^2 (e_1' q_g)^2 (e_1' q_f)^2 \\
&\leq \frac{N}{\pi^2} \sum_{g=1}^{G} \pi_g^3 \delta(x_g)^4 (e_1' q_g)^4 = o(\mathbb{E}_N(\mathcal{C}_{16})).
\end{aligned}$$

It therefore follows that

$$\mathcal{C}_1 = (1 + o_{P_N}(1))\mathbb{E}_N(\mathcal{C}_1) = (1 + o_{P_N}(1)) \left(e_1' Q_N^{-1} \Omega Q_N^{-1} e_1 + \frac{N-1}{\pi} \sum_{g=1}^{G} (e_1' q_g)^2 \pi_g^2 \delta(x_g)^2 \right).$$

Finally, the cross-term $\mathcal{C}_3$ is $o_{P_N}(\mathbb{E}_N(\mathcal{C}_1)^{1/2})$ by Cauchy-Schwarz inequality, so that $\widehat{\sigma}^2_{\text{CRV}} = (1 + o_{P_N}(1))\mathbb{E}_N(\mathcal{C}_1)$, which yields the result.

S2. PROOFS OF RESULTS IN SECTION IV

For the proof of Propositions 1, we suppose that Assumptions 1 and 2 (i) hold. We denote the conditional variance of τ by $\tilde{\sigma}_\tau^2 = e_1' H Q_N^{-1} H \tilde{\Omega} H Q_N^{-1} H e_1$, where $\tilde{\Omega} = \mathbb{E}[\sigma^2(X_i)\lambda(X_i)]/\pi$, and $\lambda(x) = m(x/h) \cdot m(x/h)' \mathbb{I}\{|x| \leq h\}$. We assume that $\tilde{\sigma}_\tau^2$ is bounded and bounded away from zero. To ensure that $\widehat{\sigma}^2_{NN}$, as defined in Section IV.A, is consistent, we assume that as $N \to \infty$, $P_N(\sum_{g=1}^{G_h} \mathbb{I}\{n_g \leq 1\}) \to 0$, so that in large samples there are at least two observations available for each support point. We put $\widehat{\sigma}_g^2 = 0$ if $n_g \leq 1$. For simplicity, we also assume that $h \to 0$; otherwise r_{sup} will diverge with the sample size.

For the proof of Proposition 2, we suppose that Assumptions 1 and 2 hold. For simplicity, we also assume that as $N \to \infty$, G_h^+ and G_h^- are fixed, and that $\min_{g \in \mathcal{G}_h} \pi_g/\pi$ is bounded away from zero. We also assume that the asymptotic variance of $\widehat{\tau} + \widehat{b}(W)$ is bounded away from zero for some $W \in \mathcal{W}$.

S2.1. **Proof of Proposition 1**

We first derive the expression for r_{sup}, following the arguments in Theorem B.1 in Armstrong and Kolesár (2016). Note first that the local linear estimator $\widehat{\tau}$ can be written as a linear estimator, $\widehat{\tau} = \sum_{i=1}^N w(X_i)Y_i$, with the weights $w(x)$ given in (8). Put $w_+(x) = w(x)\mathbb{I}\{x \geq 0\}$, and $w_-(x) = w(x)\mathbb{I}\{x < 0\}$, and put $\mu_+(x) = \mu(x)\mathbb{I}\{x \geq 0\}$ and $\mu_-(x) = \mu(x)\mathbb{I}\{x < 0\}$, with the convention that $\mu_-(0) = \lim_{x \uparrow 0} \mu(0)$. Since $\sum_{i=1}^N w_+(X_i) = -\sum_{i=1}^N w_-(X_i) = 1$ and $\tau = \mu_+(0) - \mu_-(0)$, the conditional bias has the form

$$\tilde{\tau}_h - \tau = \sum_i w_+(X_i)(\mu_+(X_i) - \mu_+(0)) + \sum_i w_-(X_i)(\mu_-(X_i) - \mu_-(0)).$$

By assumption, the first derivatives of the functions μ_+ and μ_- are Lipschitz, and hence absolutely continuous, so that, by the Fundamental Theorem of Calculus and Fubini's theorem, we can write, for $x \geq 0$, $\mu_+(x) = \mu_+(0) + \mu_+'(0)x + r(x)$, and for $x < 0$, $\mu_-(x) = \mu_-(0) + \mu_-'(0)x + r(x)$, where $r(x) = \mathbb{I}\{x \geq 0\}\int_0^x \mu''(s)(x-s)\,ds + \mathbb{I}\{x < 0\}\int_x^0 \mu''(s)(x-s)\,ds$. Since the weights satisfy $\sum_{i=1}^N X_i w_+(X_i) = 0$, and $\sum_{i=1}^N X_i w_-(X_i) = 0$, it follows that

$$\begin{aligned}
\tilde{\tau}_h - \tau &= \sum_{i\colon X_i \geq 0} w(X_i)r(X_i) + \sum_{i\colon X_i < 0} w(X_i)r(X_i) \\
&= \int_0^\infty \mu''(s) \sum_{i\colon X_i \geq s} w(X_i)(X_i - s)\,ds + \int_{-\infty}^0 \mu''(s) \sum_{i\colon X_i \leq -s} w(X_i)(X_i - s)\,ds,
\end{aligned}$$

where the second line uses Fubini's theorem to change the order of summation and integration. Next, note that $\bar{w}_+(s) = \sum_{i\colon X_i \geq s} w(X_i)(X_i - s)$ is negative for all $s \geq 0$, because $\bar{w}_+(0) = 0$, $\bar{w}_+(s) = 0$ for $s \geq h$, and $\bar{w}_+'(s) = -\sum_{X_i \geq s} w(X_i)$ is monotone on $[0, h]$ with $\bar{w}_+'(0) = -1$. Similarly, $\bar{w}_-(s) = \sum_{i\colon X_i \leq -s} w(X_i)(X_i - s)$ is positive for all $s \geq 0$. Therefore, the expression in the preceding display is maximized by setting $\mu''(x) = -K\operatorname{sign}(x)$, and minimized by setting $\mu''(x) = K\operatorname{sign}(x)$. Plugging these expressions into the preceding display then gives $|\tilde{\tau}_h - \tau| \leq B_N$, with $B_N = -K\sum_{i=1}^N w(X_i)X_i^2 \operatorname{sign}(X_i)/2$, which yields (8).

Let $o_{P_N}(1)$ denote a term that's asymptotically negligible, uniformly over $\mathcal{M}_{\text{H}}(K)$. To complete the proof, we need to show that (i) $\widehat{\sigma}_{NN} = \tilde{\sigma}_\tau^2 + o_{P_N}(1)$, (ii) $\sqrt{N_h}(\widehat{\tau} - \tilde{\tau}_h) = \mathcal{N}(0, \tilde{\sigma}_\tau^2) + o_{P_N}(1)$, (iii) $\sqrt{N_h}(\tilde{\tau}_h - \tau) = b(\mu) + o_{P_N}(1)$, and (iv) $\sqrt{N_h}B_N = B_\infty + o_{P_N}(1)$, where $b(\mu) = \sqrt{N/\pi} \cdot e_1' Q_N^{-1}\mathbb{E}[M_i r(X_i)]$ and $B_\infty = -(K/2\sqrt{N/\pi}) \cdot e_1' Q_N^{-1}\mathbb{E}[M_i X_i^2 \operatorname{sign}(X_i)]$ are non-random, and by an argument analogous to that in the preceding paragraph, satisfy $\sup_{\mu \in \mathcal{M}_{\text{H}}(K)} |b(\mu)| \leq B_\infty$. It then follows from uniform continuity of $\text{cv}_{1-\alpha}(\cdot)$ that

$$P_N(\sqrt{N_h}|\widehat{\tau} - \tau| \leq \text{cv}_{1-\alpha}(r_{\text{sup}})\widehat{\sigma}_{NN}^2) = P_N(|Z + b(\mu)/\tilde{\sigma}_\tau^2| \leq \text{cv}_{1-\alpha}(B_\infty/\tilde{\sigma}_\tau^2)\tilde{\sigma}_\tau^2) + o_{P_N}(1),$$

where $Z \sim \mathcal{N}(0, 1)$, from which honesty follows.

To show (i), note that by (S1), (S2), and the law of large numbers, it suffices to show that $H^{-1}\widehat{\Omega}_{NN}H^{-1}-N_h^{-1}\sum_{i=1}^{N}\sigma^2(X_i)\lambda(X_i)=o_{P_N}(1)$. Note that $N_h\widehat{\Omega}_{NN}=\sum_i \mathbb{I}\{n(X_i)>1\}\cdot \epsilon_i^2\lambda(X_i)-\sum_{i>j}\epsilon_i\epsilon_j\mathbb{I}\{X_i=X_j, n(X_i)>1\}/(n(X_i)-1)$, where $n(x)=\sum_{j=1}^{N}\mathbb{I}\{X_j=x\}$ (so that $n(x_g)=n_g$). This yields the decomposition

$$
\begin{aligned}
H^{-1}\widehat{\Omega}_{NN}H^{-1}-\frac{1}{N_h}\sum_{i=1}^{N}\sigma^2(X_i)\lambda(X_i) &= \frac{1}{N_h}\sum_{i=1}^{N}(\epsilon_i^2-\sigma^2(X_i))\lambda(X_i)\\
&+\frac{(1+o_{P_N}(1))\cdot 2}{\sqrt{N\pi N_h}}\sum_{i>j}\epsilon_i\epsilon_j\frac{\mathbb{I}\{n(X_i)>1\}\mathbb{I}\{X_i=X_j\}\lambda(X_i)}{n(X_i)-1}\\
&-\frac{1}{N_h}\sum_{i=1}^{N}\lambda(X_i)\epsilon_i^2\mathbb{I}\{n(X_i)=1\}.
\end{aligned}
$$

Since elements of $\lambda(X_i)$ are bounded by $\mathbb{I}\{|X_i|\le h\}$, the first term on the right-hand side of the preceding display is of the order $o_{P_N}(1)$ by the law of large numbers. Conditional on $X_1,\dots,X_N$, the second term has mean zero and variance bounded by $\frac{8}{N\pi}\max_i\sigma^4(X_i)$, which implies that unconditionally, it also has mean zero, and variance that converges to zero. Therefore, by Markov's inequality, it is also of the order $o_{P_N}(1)$. Finally, by assumption of the proposition, the probability that the third term is zero converges to one. Thus, $\widehat{\sigma}_{NN}=\tilde{\sigma}_\tau^2+o_{P_N}(1)$ as required.

Next, (ii) holds by (S1), (S2), and a central limit theorem. To show (iii), note that

$$
\sqrt{N}_h(\tau_h-\tau)=(1+o_{P_N}(1))e_1'Q_N^{-1}H\frac{1}{\sqrt{N\pi}}\sum_i H^{-1}M_i r(X_i).
$$

Since elements of $H^{-1}M_ir(X_i)$ are bounded by $\mathbb{I}\{|X_i|\le h\}KX_i^2/2\le \mathbb{I}\{|X_i|\le h\}Kh^2/2$, it follows that elements of the variance matrix of $(N\pi)^{-1/2}\sum_i H^{-1}M_ir(X_i)$ are bounded by $K^2h^4/4$. Thus, (iii) follows by Markov's inequality. Finally, the proof of (iv) is analogous.

S2.2. **Proof of Proposition 2**

It suffices to show that for each $W\in\mathcal{W}$, the left- and right-sided CIs $[c_L^{1-\alpha}(W),\infty)$ and $(-\infty,c_R^{1-\alpha}(W)]$ are asymptotically valid CIs for $\tau_h+b(W)$, for any sequence of probability laws P_N satisfying the assumptions stated at the beginning of Section S2, and satisfying $\mu\in\mathcal{M}_{\text{BME}}(h)$. Honesty will then follow by the union-intersection principle and the definition of $\mathcal{M}_{\text{BME}}(h)$.

Note first that by the central limit theorem and the delta method,

$$
\sqrt{N_h}\begin{pmatrix}\operatorname{vec}(\widehat{\mu}_g-\mu_g)\\ N_h^{-1}\sum_{i=1}^{N}H^{-1}M_iu_i\end{pmatrix}=_d\mathcal{N}\left(0,\begin{pmatrix}\pi\operatorname{diag}(\sigma_g^2/\pi_g) & \operatorname{vec}(\sigma_g^2m_g'H^{-1})\\ \operatorname{vec}(\sigma_g^2m_g'H^{-1})' & H^{-1}\Omega H^{-1}\end{pmatrix}\right)+o_{P_N}(1).
$$

Applying the delta method again, along with Lemma S1, yields

$$
\sqrt{N_h}\begin{pmatrix}\operatorname{vec}(\widehat{\delta}(x_g)-\delta(x_g))\\ \widehat{\tau}_h-\tau_h\end{pmatrix}=_d\mathcal{N}(0,\Sigma)+o_{P_N}(1),
$$

where the variance matrix Σ is given by

$$\Sigma = \begin{pmatrix} \operatorname{diag}(\sigma_g^2 \cdot \pi/\pi_g) + \mathcal{V} & \operatorname{vec}(\sigma_g^2 m_g' Q_N^{-1} e_1 - m_g' Q_N^{-1} \Omega Q_N^{-1} e_1) \\ \operatorname{vec}(\sigma_g^2 m_g' Q_N^{-1} e_1 - m_g' Q_N^{-1} \Omega Q_N^{-1} e_1)' & e_1' Q_N^{-1} \Omega Q_N^{-1} e_1 \end{pmatrix},$$

and $\mathcal{V}$ is a $G_h \times G_h$ matrix with (g, g^*) element equal to $m_g' Q^{-1} \Omega Q^{-1} m_{g^*} - (\sigma_g^2 + \sigma_{g^*}^2) m_g' Q^{-1} m_{g^*}$.

Fix $W = (g^-, g^+, s^-, s^+)$, and let $a(W) \in \mathbb{R}^{G_h+1}$ denote a vector with the g_-th element equal to s^-, $(G_h^- + g_+)$th element equal to s^+, the last element equal to one, and the remaining elements equal to zero. It follows that $\sqrt{N_h}(\hat{\tau} + \hat{b}(W) - \tau_h - b(W))$ is asymptotically normal with variance $a(W)'\Sigma a(W)$. To construct the left- and right-sided CIs, we use the variance estimator

$$\hat{V}(W) = a(W)'\hat{\Sigma} a(W), \tag{S8}$$

where $\hat{\Sigma}$ is a plug-in estimator of Σ that replaces Q_N by $\hat{Q}$, Ω by $\hat{\Omega}_{\text{EHW}}$, π/π_g by N_h/n_g, and σ_g^2 by $\hat{\sigma}_g^2$ (given in Section IV.A). Since by standard arguments $n_g/N_h = \pi_g/\pi + o_{P_N}(1)$, and $\hat{\sigma}_g^2 = \sigma_g^2 + o_{P_N}(1)$, it follows from (S2) and (S5) that $\hat{V}(W) = a(W)'\Sigma a(W) + o_{P_N}(1)$, which, together with the asymptotic normality of $\sqrt{N_h}(\hat{\tau} + \hat{b}(W) - \tau_h - b(W))$, implies asymptotic validity of $[c_L^{1-\alpha}(W), \infty)$ and $(-\infty, c_R^{1-\alpha}(W)]$, as required.

S3. ADDITIONAL FIGURES

This section shows the fit of the specifications considered in Section I. Specifically, Figure S1 shows the fit of a linear specification ($p = 1$) for the four values of the bandwidth h considered; Figure S2 shows the analogous results for a quadratic fit ($p = 2$). In each case, the value of the parameter τ_h is equal to height of the jump in the fitted line at the 40-year cutoff.

S4. PERFORMANCE OF HONEST CIS AND ALTERNATIVE CRV CIS IN CPS SIMULATION STUDY

In this section, we again consider the CPS placebo study from Section I in the main text to study the performance of the honest CIs proposed in the Section IV. Given that the typical increase in log wages is about 0.017 per extra year of age, guided by the heuristic in Section IV.A, we set $K = 0.045$ for the BSD CIs. In addition, we also consider two modifications of the CRV CIs that have been shown to perform better in settings with a few clusters. The first modification, which we term CRV2, is a bias-reduction modification analogous to the HC2 modification of the EHW variance estimator developed by Bell and McCaffrey (2002). This modification makes the variance estimator unbiased when the errors $u_i = Y_i - M_i'\theta_h$ are homoscedastic and independent. To describe it, let $\mathbf{M}$ denote the $N_h \times (2p+2)$ design matrix with ith row given by M_i', and let $\mathbf{M}_g$ denote the n_g rows of the design matrix $\mathbf{M}$ for which $M_i' = m(x_g)'$ (that is, those rows that correspond to "cluster" g). Here n_g is the number of observations with $X_i = x_g$. The CRV2 variance estimator replaces $\hat{\Omega}_{\text{CRV}}$ in the definition of $\hat{\sigma}^2_{\text{CRV}}$ given in Section II.A with

$$\hat{\Omega}_{\text{CRV2}} = \frac{1}{N_h} \sum_{g=1}^{G_h} \mathbf{M}_g' \mathbf{A}_g \hat{\mathbf{u}}_g \hat{\mathbf{u}}_g' \mathbf{A}_g \mathbf{M}_g,$$

where $\hat{\mathbf{u}}_g$ is an n_g-vector of the regression residuals corresponding to "cluster" g, and $\mathbf{A}_g$ is the inverse of the symmetric square root of the $n_g \times n_g$ block of the annihilator matrix $\mathbf{I}_{N_h} - \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'$ corresponding to "cluster" g, $\mathbf{I}_{n_g} - \mathbf{M}_g(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}_g'$. Thus, $\mathbf{A}_g[\mathbf{I}_{n_g} - \mathbf{M}_g(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}_g']\mathbf{A}_g = \mathbf{I}_{n_g}$. In contrast, the CRV estimator, as implemented in STATA, sets $\mathbf{A}_g = \sqrt{G_h/(G_h-1) \times (N_h-1)/(N_h-2(p+1))}\mathbf{I}_{n_g}$. The second modification that we consider also uses the CRV2 variance estimator, but replaces the usual 1.96 critical value in the construction of 95% confidence intervals with a critical value based on a t-distribution with a degrees of freedom, where a is chosen so that the first two moments of $\widehat{\sigma}_{\text{CRV2}}/\mathbb{V}(\widehat{\tau} \mid \mathbf{M})$ match that of χ^2_a/a, assuming correct specification and independent homoscedastic errors. This modification has also been proposed by Bell and McCaffrey (2002), and, as discussed in Imbens and Kolesár (2016), it generalizes the classic Welch-Sattherthwaite solution to the Behrens-Fisher problem.

The results are reported in Table S1 for the linear specification ($p = 1$), and in Table S2 for the quadratic specification ($p = 2$). For convenience, the tables also reproduce the results for inference based on EHW and CRV standard errors, reported in Table 1 in the main text. To compare the length of honest CIs and CRV-BM CIs to those of EHW, CRV, and CRV2 CIs, the tables report average normalized standard errors. We define normalized standard error of a CI $[a, b]$ with nominal level 95% as $(b - a)/(2 \times 1.96)$, so that the CI is given by adding and subtracting the normalized standard error times the usual 1.96 critical value from its midpoint. Finally, Table S3 compares magnitudes of the various clustered standard errors relative to that of the conventional EHW estimator.

The coverage properties of honest CIs are excellent: they achieve at least 95% coverage in all specifications. The length of BME CIs is moderately larger than that of EHW CIs in specifications in which EHW CIs achieve proper coverage. The BME CIs are much longer, especially for large h or small N_h. This is in line with the discussion in Section IV.B.

In designs in which the fitted model is close to being correctly specified (all quadratic specifications and linear specifications with $h < 15$) CRV-BM CIs do about as well as EHW CIs in delivering approximately correct coverage, while the CRV2 adjustment alone is not sufficient to bring coverage close to 95%. On the other hand, in designs in which the fitted model is misspecified (linear specifications with $h = 15$ or $h = \infty$), the coverage of CRV-BM CIs is only slightly better than that of EHW CIs; the improvement in coverage is not sufficient to bring coverage close to 95%. This contrasts with the coverage of the honest CIs, which remains close to, or above, 95%. Furthermore, the CRV-BM CIs are much more variable than EHW, and thus the average length of the CRV-BM CI, as measured by the normalized SE, is greater than in the case of EHW: in the designs considered, they are about 40% longer on average, sometimes more than twice as long. In part due to this variability, the probability that CRV2 and CRV-BM CIs are longer than EHW CIs in a particular sample is not very close to 1, and falls between 0.7 and 0.9 for most of the designs that we consider, as shown in Table S3. This suggests, in line with the results in Section V, that in any particular empirical application, these CIs may not be larger than EHW CIs.

S5. ADDITIONAL SIMULATION EVIDENCE

In this section, we consider a second Monte Carlo exercise in which we simulate realizations of an outcome variable Y_i and a running variable X_i from several data generating processes (DGPs) with different conditional expectation functions and different numbers of support points, and also several sample sizes. This allows us to disentangle the effect of model misspecification and the number of the support points on the performance of CRV-based inference; something that was not possible in the CPS placebo study from Section I.

Each of our DGPs is such that the support of the running variable is the union of an equally spaced grid of G^- points $-1, -(G^- - 1)/G^-, \ldots, -1/G^-$ and an equally spaced grid of G^+ points $1/G^+, 2/G^+, \ldots, 1$. We consider values $G^-, G^+ \in \{5, 25, 50\}$. The distribution of X_i then has probability mass 1/2 spread equally across the support points above and below zero, so that $P(X_i = x_g) = 1/G^+$ for $x_g > 0$ and $P(X_i = x_g) = 1/G^-$ for $x_g < 0$.

The outcome variable is generated as $Y_i = \mu(X_i) + \varepsilon_i$, where ε_i and X_i are independent, $\varepsilon_i \sim \mathcal{N}(0, 0.1)$, and

$$\mu(x) = x + \lambda_1 \cdot \sin(\pi \cdot x) + \lambda_2 \cdot \cos(\pi \cdot x).$$

We also generate a treatment indicator that is equal to one if $X_i \geq 0$, and equal to zero otherwise. Since $\mu(x)$ is continuous at $x = 0$ for every (λ_1, λ_2), the causal effect of our treatment at the cutoff is zero in our all our DGPs. We consider $(\lambda_1, \lambda_2) \in \{(0, 0), (0.05, 0), (0, 0.05)\}$ and the sample sizes $N_h \in \{100, 1000, 10000\}$, and estimate the treatment effect by fitting the linear model

$$Y_i = \beta_0 + \tau \cdot \mathbb{I}\{X_i \geq 0\} + \beta^- \cdot X_i + \beta^+ \cdot \mathbb{I}\{X_i \geq 0\} \cdot X_i + U_i. \tag{S9}$$

Note that this specification is analogous to the model (1) in the main text with $p = 1$ and $h = 1$. We do not consider alternative values of h and p in this simulation exercise because variation in the accuracy of the fitted model is achieved by varying the DGP.

To asses the accuracy of model (S9), we plot the versions of $\mu(x)$ that we consider together with the corresponding linear fit in Figure S3 for the case that $G^- = G^+ = 10$. As one can see, the departure from linearity is rather modest for $(\lambda_1, \lambda_2) \in \{(0.05, 0), (0, 0.05)\}$. In Tables S4–S6 we then report the empirical standard deviation of $\widehat{\tau}$, the empirical coverage probabilities of the EHW and CRV CIs with nominal level 95%, as well as the coverage probabilities of the honest CIs, and the CRV2 and CRV-BM modifications. For BSD CIs, we consider the values $K = \pi^2/20$, which corresponds to the largest value of the second derivative of μ for the designs considered, as well as the more conservative choice $K = \pi^2/10$.[1] To compare length, the tables report normalized standard errors, defined in Section S4.

Table S4 reports results for the case $(\lambda_1, \lambda_2) = (0, 0)$, in which the true conditional expectation function is linear and thus our fitted model is correctly specified. We see that the CRV standard error is a downward-biased estimate of the standard deviation of $\widehat{\tau}$, and therefore the CRV confidence interval under-covers the treatment effect. The distortion is

[1]As pointed out in the main text, the choice of K requires subject knowledge, as it implies certain restrictions on the shape of $\mu(x)$. This is generally difficult to mimic in the context of a simulation study based on an artificial DGP.

most severe for the case with the least number of points on either side of the threshold ($G^- = G^+ = 5$), where it amounts to a deviation of 20 percentage points from the nominal level. With more support points the distortion becomes less pronounced, but it is still noticeable even for $G^- = G^+ = 50$. These findings are the same for all the sample sizes we consider. The distortion of CRV2 CIs is slightly smaller, and the coverage of CRV-BM is close to 95%, at the cost of a loss in power (as measured by the average normalized standard error). The honest CIs perform well in terms of coverage, although they are quite conservative: this is the price for maintaining good coverage over CEFs that are less smooth than $\mu(x) = x$ (see Armstrong and Kolesár (2017) for theoretical results on impossibility of adaptation to smooth functions).

Table S5 reports results for the case $(\lambda_1, \lambda_2) = (0, .05)$. Here $\mu(x)$ is nonlinear, but due to the symmetry properties of the cosine function $\tau_h = 0$. This setup mimics applications in which the bias of $\widehat{\tau}$ is small even though the functional form of $\mu(x)$ is misspecified. In line with our asymptotic approximations, the CRV standard error is downward biased for smaller values of N, and upward biased for larger sample sizes. Simulation results for the case that $N = 10^7$, which are not reported here, also confirm that the CRV standard error does not converge to zero. Correspondingly, the CRV CI under-covers the treatment effect for smaller values of N, and over-covers for larger values. The distortions are again more pronounced for smaller values of G^+ and G^-. The CRV-BM CIs correct for these size distortions, while the CRV2 modification alone is not sufficient. The honest CIs perform well in terms of coverage, but they are again quite conservative: again this is due to the possible bias adjustment built into the CIs.

Table S6 reports results for the case $(\lambda_1, \lambda_2) = (.05, 0)$. Here the linear model is misspecified as well, but in such a way that τ_h is substantially different from zero; with its exact value depending on G^+ and G^-. As with the previous sets of results, the CRV standard error is downward biased for smaller values of N, and upward biased for larger sample sizes. However, since $\tau_h \neq 0$ here, the coverage probability of the CRV confidence interval is below the nominal level for all N, and tends to zero as the sample size increases. For smaller values of N, the coverage properties of the CRV confidence interval are also worse than those of the standard EHW confidence interval. The CRV confidence interval only performs better in a relative sense than EHW confidence interval when N is large, but for these cases both CIs are heavily distorted and have coverage probability very close to zero. So in absolute terms the performance of the CRV confidence interval is still poor. The same conclusion applies to the CRV2 and CRV-BM modifications. The assumption underlying the construction of BME CIs is violated for $G_+ = G_- = 5$, since in this case, the specification bias of a linear approximation to the CEF is actually worse at zero than at other support points. Consequently, BME CIs undercover once N_h is sufficiently large. In contrast, the coverage of BSD CIs remains excellent for all specifications, as predicted by the theory.

In summary, the simulation results for EHW and CRV CIs are in line with the theory presented in Sections II and III in the main text, and the results for CRV2 and CRV-BM CIs are in line with the results presented in Section S4. The honest CIs perform well in terms of coverage, although BME CIs are overly conservative if N_h is small or the number of support points is large. As discussed in Section IV.B, they are best suited to applications with a few support points, in with a sufficient number of observations available for each support point. BSD CIs, combined with appropriate choice of the smoothness constant K, perform very well.

S6. LOWER BOUND FOR K

In this section, we use ideas in Armstrong and Kolesár (2017) to derive and estimate a left-sided CI for a lower bound on the smoothness constant K under the setup of Section IV.A in the main text.

To motivate our procedure, consider measuring the curvature of the CEF μ over some interval $[x_1, x_3]$ by measuring how much it differs from a straight line at some point $x_2 \in (x_1, x_3)$. If the function were linear, then its value at x_2 would equal $\lambda\mu(x_1) + (1-\lambda)\mu(x_3)$, where $\lambda = (x_3 - x_2)/(x_3 - x_1)$ is the distance of x_3 to x_2 relative to the distance to x_1. The next lemma gives a lower bound on K based on the curvature estimate $\lambda\mu(x_1) + (1-\lambda)\mu(x_3) - \mu(x_2)$.

Lemma S2. *Suppose $\mu \in \mathcal{M}_H(K)$ for some $K \geq 0$. Then $K \geq |\Delta(x_1, x_2, x_3)|$, where*

$$\Delta(x_1, x_2, x_3) = 2\frac{(1-\lambda)\mu(x_3) + \lambda\mu(x_1) - \mu(x_2)}{(1-\lambda)x_3^2 + \lambda x_1^2 - x_2^2} = 2\frac{(1-\lambda)\mu(x_3) + \lambda\mu(x_1) - \mu(x_2)}{(1-\lambda)\lambda(x_3 - x_1)^2}. \quad \text{(S10)}$$

Proof. We can write any $\mu \in \mathcal{M}_{\mathrm{H}}(K)$ as

$$\mu(x) = \mu(0) + \mu'(0)x + r(x) \qquad r(x) = \int_0^x (x-u)\mu''(u)\,\mathrm{d}u. \quad \text{(S11)}$$

This follows by an argument given in the proof of Proposition 1 in Section S2. Hence,

$$\begin{aligned}\lambda\mu(x_1) + (1-\lambda)\mu(x_3) - \mu(x_2) &= (1-\lambda)r(x_3) + \lambda r(x_1) - r(x_2) \\ &= (1-\lambda)\int_{x_2}^{x_3}(x_3 - u)\mu''(u)\,\mathrm{d}u + \lambda\int_{x_1}^{x_2}(u - x_1)\mu''(u)\,\mathrm{d}u.\end{aligned}$$

The absolute value of the right-hand side is bounded by K times $(1-\lambda)\int_{x_2}^{x_3}(x_3 - u)\,\mathrm{d}u +$ $+\lambda\int_{x_1}^{x_2}(u - x_1)\,\mathrm{d}u = (1-\lambda)(x_3 - x_2)^2/2 + \lambda(x_2 - x_1)^2/2$, which, combined with the definition of λ, yields the result. □

Remark S1. The maximum departure of μ from a straight line over $[x_1, x_3]$ is given by $\max_{\lambda\in[0,1]}|(1-\lambda)\mu(x_3) + \lambda\mu(x_1) - \mu(\lambda x_1 + (1-\lambda)x_3)|$. Lemma S2 implies that this quantity is bounded by $\max_{\lambda\in[0,1]}(1-\lambda)\lambda(x_3 - x_1)^2 K/2 = K(x_3 - x_1)^2/8$, with the maximum achieved at $\lambda = 1/2$. We use this bound for the heuristic for the choice of K described in Section IV.A.

While it is possible to estimate a lower bound on K using Lemma S2 by fixing points x_1, x_2, x_3 in the support of the running variable and estimating $\Delta(x_1, x_2, x_3)$ by replacing $\mu(x_j)$ in (S10) with the sample average of Y_i for observations with $X_i = x_j$, such a lower bound estimate may be too noisy if only a few observations are available at each support point. To overcome this problem, we consider curvature estimates that average the value of μ over s neighboring support points. To that end, the following generalization of Lemma S2 will be useful.

Lemma S3. *Let $I_k = [\underline{x}_k, \bar{x}_k]$, $k = 1, 2, 3$ denote three intervals on the same side of cutoff such that $\bar{x}_k \leq \underline{x}_{k+1}$. For any function g, let $E_n[g(x_k)] = \sum_{i=1}^{N} g(x_i)\mathbb{I}\{x_i \in I_k\}/n_k$, $n_k = \sum_{i=1}^{N} \mathbb{I}\{x_i \in I_k\}$ denote the average value in interval k. Let $\lambda_n = E_n(x_3 - x_2)/E_n(x_3 - x_1)$. Then $K \geq |\mu(I_1, I_2, I_3)|$, where*

$$\Delta(I_1, I_2, I_3) = 2\frac{\lambda_n E_n\mu(x_1) + (1-\lambda_n)E_n\mu(x_3) - E_n\mu(x_2)}{E_n\left[(1-\lambda_n)x_3^2 + \lambda_n x_1^2 - x_2^2\right]}. \tag{S12}$$

Proof. By Equation (S11), for any $\lambda \in [0, 1]$ and any three points x_1, x_2, x_3, we have

$$\begin{aligned}\lambda\mu(x_1) + (1-\lambda)\mu(x_3) - \mu(x_2) &= \delta\mu'(0) + \lambda r(x_1) + (1-\lambda)r(x_3) - r(x_2) \\ &= (1-\lambda)\int_{x_2}^{x_3}(x_3 - u)\mu''(u)\,\mathrm{d}u + \lambda\int_{x_1}^{x_2}(u - x_1)\mu''(u)\,\mathrm{d}u + \delta\left(\mu'(0) + \int_0^{x_2}\mu''(u)\,\mathrm{d}u\right)\end{aligned}$$

where $\delta = (1-\lambda)x_3 + \lambda x_1 - x_2$. Setting $\lambda = \lambda_n$, taking expectations with respect to the empirical distribution, observing that $E_n\delta = 0$ and $E_n x_2 = (1-\lambda_n)E_n x_3 + \lambda_n E_n x_1$, and using iterated expectations yields

$$\begin{aligned}&\lambda_n E_n\mu(x_1) + (1-\lambda_n)E_n\mu(x_3) - E_n\mu(x_2) \\ &\qquad = (1-\lambda_n)E_n\int_{\bar{x}_2}^{x_3}(x_3 - u)\mu''(u)\,\mathrm{d}u + \lambda_n E_n\int_{x_1}^{\underline{x}_2}(u - x_1)\mu''(u)\,\mathrm{d}u + S,\end{aligned} \tag{S13}$$

where

$$\begin{aligned}S &= (1-\lambda_n)E_n\int_{x_2}^{\bar{x}_2}(E_n x_3 - u)\mu''(u)\,\mathrm{d}u + E_n\int_{\underline{x}_2}^{x_2}(\lambda_n u - \lambda_n E_n x_1 + E_n x_2 - x_2)\mu''(u)\,\mathrm{d}u \\ &= \int_{\underline{x}_2}^{\bar{x}_2} q(u)\mu''(u)\,\mathrm{d}u, \qquad q(u) = (1-\lambda_n)(E_n x_3 - u) + E_n\mathbb{I}\{u \leq x_2\}(u - x_2).\end{aligned}$$

Observe that the function $q(x)$ is weakly positive on I_2, since $q(\underline{x}_2) = (1-\lambda_n)(E_n x_3 - \underline{x}_2) + \underline{x}_2 - E_n x_2 = \lambda_n(\underline{x}_2 - E_n x_1) \geq 0$, $q(\bar{x}_2) = (1-\lambda_n)(E_n x_3 - \bar{x}_2) \geq 0$, and $q'(u) = \lambda_n - E_n\mathbb{I}\{x_2 \leq u\}$, so that the first derivative is positive at first and changes sign only once. Therefore, the right-hand side of (S13) is maximized over $\mu \in \mathcal{F}_{\mathrm{M}}(K)$ by taking $\mu''(u) = K$ and minimized by taking $\mu''(u) = -K$, which yields

$$|\lambda_n E_n\mu(x_1) + (1-\lambda_n)E_n\mu(x_3) - E_n\mu(x_2)| \leq \frac{K}{2}E_n\left[(1-\lambda_n)x_3^2 + \lambda_n x_1^2 - x_2^2\right].$$

The result follows. □

To estimate a lower bound on K for three fixed intervals I_1, I_2 and I_3, we simply replace $E_n\mu(x_k)$ in (S12) with the sample average of Y_i over the interval I_k. We do this over multiple interval choices. To describe our choice of intervals, denote the support points of the running variable by $x^-_{G_-} < \cdots < x^-_1 < 0 \leq x^+_1 < \cdots < x^+_{G_+}$, and let $J^+_{ms} = [x^+_{ms+1-s}, x^+_{ms}]$ denote the mth interval containing s support points closest to the threshold, and similarly let $J^-_{ms} = [x^-_{ms}, x^-_{ms+1-s}]$. To estimate the curvature based on the intervals $J^+_{3k-2,s}$, $J^+_{3k-1,s}$ and $J^+_{3k,s}$, we use the plug-in estimate

$$\hat{\Delta}^+_{ks} = 2\frac{\lambda_{ks}\bar{y}(J^+_{3k-2,s}) + (1-\lambda_{ks})\bar{y}(J^+_{3k-1,s}) - \bar{y}(J^+_{3k,s})}{(1-\lambda_{ks})\bar{x}^2(J^+_{3k,s}) + \lambda_{ks}\bar{x}^2(J_{3k-2,s}) - \bar{x}^2(J^+_{3k-1,s})},$$

where we use the notation $\bar{y}(I) = \sum_{i=1}^N Y_i \mathbb{I}\{X_i \in I\}/n(I)$, $n(I) = \sum_{i=1}^N \mathbb{I}\{X_i \in I\}$, and $\bar{x}^2(I) = \sum_{i=1}^N X_i \mathbb{I}\{X_i \in I\}/n(I)$, and $\lambda_{ks} = (\bar{x}(J^+_{3k,s}) - \bar{x}(J^+_{3k-1,s}))/(\bar{x}(J^+_{3k,s}) - \bar{x}(J^+_{3k-2,s}))$.

For simplicity, we assume that the regression errors $\epsilon_i = Y_i - \mu(X_i)$ are normally distributed with conditional variance $\sigma^2(X_i)$, conditionally on the running variable. Then $\hat{\Delta}^+_{ks}$ is normally distributed with variance

$$\mathbb{V}(\hat{\Delta}^+_{ks}) = 2\frac{\lambda^2_{ks}\bar{\sigma}^2(J^+_{3k-2,s})/n(J^+_{3k-2,s}) + (1-\lambda_{ks})\bar{\sigma}^2(J^+_{3k-1,s})/n(J^+_{3k-1,s}) + \bar{\sigma}^2(J^+_{3k,s})/n(J^+_{3k,s})}{\left[(1-\lambda_{ks})\bar{x}^2(J^+_{3k,s}) + \lambda_{ks}\bar{x}^2(J_{3k-2,s}) - \bar{x}^2(J^+_{3k-1,s})\right]^2}.$$

By Lemma S3 the mean of this curvature estimate is bounded by $|E[\hat{\Delta}^+_{ks}]| \leq K$. Define $\hat{\Delta}^-_{ks}$ analogously.

Our construction of a lower bound for K is based on the vector of estimates $\hat{\Delta}_s = (\hat{\Delta}^-_{1s}, \ldots, \hat{\Delta}^-_{M_-s}, \hat{\Delta}^+_{1s}, \ldots, \hat{\Delta}^+_{M_+s})$, where M_+ and M_- are the largest values of m such that $x^+_{3m,s} \leq x_{G_+}$ and $x^-_{3m,s} \geq x_{G_-}$.

Since elements of $\hat{\Delta}_s$ are independent with means bounded in absolute value by K, we can construct a left-sided CI for K by inverting tests of the hypotheses $H_0\colon K \leq K_0$ vs $H_1\colon K > K_0$ based on the sup-t statistic $\max_{k,q\in\{-,+\}}|\hat{\Delta}^q_{ks}/\mathbb{V}(\hat{\Delta}^q_{ks})^{1/2}|$. The critical value $q_{1-\alpha}(K_0)$ for these tests is increasing in K_0 and corresponds to the $1-\alpha$ quantile of $\max_{k,q\in\{-,+\}}|Z^q_{ks} + K_0/\mathbb{V}(\hat{\Delta}^q_{ks})^{1/2}|$, where Z^q_{ks} are standard normal random variables, which can easily be simulated. Inverting these tests leads to the left-sided CI of the form $[\hat{K}_{1-\alpha}, \infty)$, where $\hat{K}_{1-\alpha}$ solves $\max_{k,q\in\{-,+\}}|\hat{\Delta}^q_{ks}/\mathbb{V}(\hat{\Delta}^q_{ks})^{1/2}| = q_{1-\alpha}(\hat{K}_{1-\alpha})$. Following Chernozhukov et al. (2013), we use $\hat{K}_{1/2}$ as a point estimate, which is half-median unbiased in the sense that $P(\hat{K}_{1/2} \leq K) \geq 1/2$.

An attractive feature of this method is that the same construction can be used when the running variable is continuously distributed. To implement this method, one has to choose the number of support points s to average over. We leave the optimal choice of s to future research, and simply use $s = 2$ in the empirical applications in Section V of the main text.

REFERENCES

Armstrong, T. B. and M. Kolesár (2016): "Simple and honest confidence intervals in nonparametric regression," ArXiv:1606.01200.

——— (2017): "Optimal inference in a class of regression models," *Econometrica*, forthcoming.

Bell, R. M. and D. F. McCaffrey (2002): "Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Smples," *Survey Methodology*, 28, 169–181.

Chernozhukov, V., S. Lee, and A. Rosen (2013): "Intersection Bounds: estimation and inference," *Econometrica*, 81, 667–737.

Imbens, G. W. and M. Kolesár (2016): "Robust Standard Errors in Small Samples: Some Practical Advice," *Review of Economics and Statistics*, 98, 701–712.

Table S1: Inference in CPS simulation study for placebo treatment: Linear specification.

				Avg. normalized SE						CI coverage rate (%)					
h	τ_h	N_h	$SD(\widehat{\tau})$	EHW	CRV	CRV2	BM	BME	BSD	EHW	CRV	CRV2	BM	BME	BSD
5	−0.008	100	0.239	0.234	0.166	0.218	0.364	0.606	0.309	93.8	77.3	86.3	96.9	100	97.2
		500	0.104	0.104	0.073	0.097	0.160	0.253	0.170	94.7	78.0	86.8	96.7	100	97.9
		2000	0.052	0.052	0.036	0.048	0.079	0.125	0.109	94.9	77.3	86.6	96.6	100	99.7
		10000	0.021	0.023	0.015	0.020	0.033	0.055	0.064	96.1	77.2	86.6	96.9	100	99.9
10	−0.023	100	0.227	0.223	0.193	0.218	0.270	0.946	0.394	93.9	87.3	90.8	95.2	100	97.0
		500	0.099	0.099	0.086	0.097	0.117	0.385	0.215	94.4	87.6	91.1	95.2	100	98.0
		2000	0.049	0.050	0.044	0.049	0.059	0.187	0.131	93.0	86.0	89.8	94.3	100	99.1
		10000	0.021	0.022	0.021	0.024	0.029	0.086	0.077	82.9	78.1	83.8	91.0	100	99.1
15	−0.063	100	0.222	0.216	0.197	0.214	0.246	1.137	0.453	92.7	88.4	90.7	94.4	100	96.8
		500	0.095	0.096	0.089	0.096	0.108	0.507	0.242	89.9	85.3	87.9	91.7	100	98.5
		2000	0.048	0.048	0.047	0.052	0.058	0.243	0.146	73.0	71.2	75.6	81.6	100	98.7
		10000	0.020	0.021	0.028	0.030	0.034	0.118	0.087	15.3	34.8	43.9	55.8	100	98.7
∞	−0.140	100	0.208	0.205	0.196	0.208	0.229	1.281	0.513	88.6	85.6	87.8	90.9	100	96.5
		500	0.091	0.091	0.094	0.099	0.107	0.822	0.269	66.7	67.3	70.8	75.6	100	97.8
		2000	0.045	0.046	0.058	0.061	0.066	0.460	0.162	13.4	29.2	34.5	41.1	100	98.0
		10000	0.019	0.020	0.043	0.046	0.049	0.271	0.097	0.0	0.6	1.3	2.9	100	98.1

Note: Results are based on 10,000 simulation runs. BM refers to CRV-BM CIs. For BSD, $M = 0.045$. For CRV-BM, BME, and BSD, Avg. normalized SE refers to average normalized standard error, described in the text. For EHW, CRV, and CRV2, it corresponds to average standard error, averaged over the simulation runs.

Table S2: Inference in CPS simulation study for placebo treatment: Quadratic specification.

				Avg. normalized SE						CI coverage rate (%)					
h	τ_h	N_h	$SD(\widehat{\tau})$	EHW	CRV	CRV2	BM	BME	BSD	EHW	CRV	CRV2	BM	BME	BSD
5	−0.100	100	0.438	0.427	0.206	0.394	0.897	0.705	0.309	93.2	60.7	84.0	97.2	99.5	97.3
		500	0.189	0.190	0.086	0.172	0.382	0.301	0.170	94.7	59.9	83.8	97.4	99.7	97.9
		2000	0.093	0.095	0.042	0.084	0.187	0.149	0.109	95.1	58.7	83.2	97.4	99.8	99.7
		10000	0.038	0.042	0.018	0.036	0.079	0.067	0.064	96.4	59.5	84.4	97.6	99.9	99.9
10	0.008	100	0.361	0.349	0.258	0.342	0.540	1.000	0.394	93.3	79.5	88.3	97.2	100	97.0
		500	0.157	0.156	0.110	0.147	0.231	0.405	0.215	94.8	79.0	88.1	96.8	100	98.0
		2000	0.077	0.078	0.055	0.073	0.115	0.196	0.131	94.7	79.4	88.1	96.8	100	99.1
		10000	0.033	0.035	0.025	0.034	0.053	0.087	0.077	95.6	80.8	89.4	97.6	100	99.1
15	0.014	100	0.349	0.329	0.270	0.325	0.443	1.189	0.453	92.3	83.6	89.1	96.0	100	96.8
		500	0.146	0.147	0.117	0.141	0.187	0.516	0.242	94.6	85.1	89.9	95.8	100	98.5
		2000	0.073	0.073	0.058	0.070	0.092	0.243	0.146	94.6	83.9	89.7	95.7	100	98.7
		10000	0.031	0.033	0.026	0.031	0.041	0.107	0.087	93.7	82.8	88.7	95.1	100	98.7
∞	−0.001	100	0.316	0.303	0.267	0.302	0.371	1.327	0.513	93.0	87.6	90.9	95.4	100	96.5
		500	0.134	0.135	0.117	0.132	0.155	0.804	0.269	94.9	89.0	92.2	95.5	100	97.8
		2000	0.068	0.067	0.058	0.065	0.077	0.402	0.162	94.7	88.7	91.6	95.4	100	98.0
		10000	0.029	0.030	0.027	0.030	0.035	0.178	0.097	96.0	91.0	93.7	96.7	100	98.1

Note: Results are based on 10,000 simulation runs. BM refers to CRV-BM CIs. For BSD, $M = 0.045$. For CRV-BM, BME, and BSD, Avg. normalized SE refers to average normalized standard error, described in the text. For EHW, CRV, and CRV2, it corresponds to average standard error, averaged over the simulation runs.

Table S3: Inference in CPS simulation study for placebo treatment: Comparison of relative magnitude of EHW standard errors relative to CRV, CRV2 and CRV-BM standard errors under linear and quadratic specification.

			Linear			Quadratic		
			Rate EHW SE <			Rate EHW SE <		
h	τ_h	N_h	CRV	CRV2	BM	CRV	CRV2	BM
5	-0.100	100	0.14	0.38	0.82	0.03	0.38	0.87
		500	0.13	0.37	0.81	0.01	0.36	0.87
		2000	0.13	0.36	0.80	0.01	0.34	0.86
		10000	0.09	0.30	0.77	0.00	0.30	0.85
10	0.008	100	0.26	0.44	0.74	0.15	0.43	0.85
		500	0.25	0.43	0.71	0.12	0.38	0.81
		2000	0.27	0.44	0.72	0.11	0.37	0.81
		10000	0.39	0.60	0.84	0.13	0.40	0.82
15	0.014	100	0.31	0.47	0.71	0.20	0.45	0.81
		500	0.34	0.47	0.69	0.18	0.39	0.74
		2000	0.45	0.61	0.79	0.17	0.38	0.72
		10000	0.92	0.96	0.99	0.18	0.39	0.72
∞	-0.001	100	0.38	0.52	0.73	0.26	0.48	0.78
		500	0.54	0.66	0.80	0.24	0.42	0.67
		2000	0.93	0.96	0.99	0.23	0.41	0.66
		10000	1.00	1.00	1.00	0.26	0.44	0.70

Note: Results are based on 10,000 simulation runs. BM refers to CRV-BM CIs. For CRV-BM CIs, SE refers to average normalized standard error, described in the text.

Table S4: Simulation results in second Monte Carlo exercise for $\mu(x) = x$.

G_+	G_-	N_h	τ_h	$sd(\hat{\tau})$	Avg. normalized SE							CI coverage rate (%)						
					EHW	CRV	CRV2	BM	BME	$BSD_{.49}$	$BSD_{.99}$	EHW	CRV	CRV2	BM	BME	$BSD_{.49}$	$BSD_{.99}$
5	5	100	0	0.148	0.149	0.105	0.141	0.221	0.332	0.197	0.249	94.6	79.1	88.0	96.6	100.0	98.5	99.0
		1000		0.046	0.047	0.032	0.044	0.068	0.097	0.091	0.124	95.7	77.5	88.0	96.6	100.0	99.0	99.5
		10000		0.015	0.015	0.010	0.014	0.022	0.031	0.047	0.067	94.8	77.3	87.4	96.2	100.0	99.5	100.0
5	25	100	0	0.143	0.141	0.114	0.136	0.186	0.686	0.178	0.219	94.5	85.4	90.3	96.9	100.0	97.8	97.8
		1000		0.044	0.044	0.036	0.043	0.057	0.248	0.078	0.102	95.5	87.2	91.1	96.9	100.0	98.6	98.1
		10000		0.014	0.014	0.011	0.013	0.018	0.074	0.037	0.052	95.5	86.7	91.1	97.2	100.0	99.4	100.0
5	50	100	0	0.143	0.140	0.115	0.135	0.183	0.754	0.176	0.216	93.7	86.0	90.3	96.3	100.0	97.4	97.8
		1000		0.043	0.044	0.036	0.042	0.056	0.397	0.077	0.100	95.3	87.3	91.3	97.1	100.0	99.0	98.1
		10000		0.014	0.014	0.011	0.013	0.018	0.111	0.036	0.050	95.2	87.3	90.9	96.6	100.0	99.2	99.9
25	25	100	0	0.136	0.131	0.124	0.131	0.143	0.791	0.159	0.186	93.8	91.3	92.9	94.8	100.0	96.6	96.9
		100000		0.040	0.041	0.039	0.041	0.044	0.280	0.064	0.076	95.5	93.2	94.5	95.7	100.0	97.9	98.0
		10000		0.013	0.013	0.012	0.013	0.014	0.083	0.027	0.032	95.4	93.1	94.3	95.5	100.0	97.8	98.3
25	50	100	0	0.132	0.130	0.125	0.130	0.140	0.843	0.156	0.183	94.2	92.8	93.8	95.4	100.0	96.9	96.7
		1000		0.042	0.041	0.039	0.041	0.043	0.430	0.063	0.074	94.7	93.0	93.9	95.0	100.0	97.4	97.6
		10000		0.013	0.013	0.012	0.013	0.013	0.119	0.026	0.031	94.9	93.2	93.9	94.8	100.0	97.9	97.8
50	50	100	0	0.133	0.129	0.126	0.130	0.138	0.879	0.154	0.180	93.5	92.6	93.2	94.7	100.0	96.6	96.7
		1000		0.040	0.041	0.040	0.041	0.042	0.472	0.062	0.072	95.3	94.1	94.8	95.4	100.0	97.6	97.7
		10000		0.013	0.013	0.012	0.013	0.013	0.127	0.025	0.029	94.9	93.9	94.3	94.9	100.0	97.5	97.7

Note: Results are based on 5,000 simulation runs. $BSD_{.49}$ and $BSD_{.99}$ refer to BSD CIs with $K = 0.493$ and $K = 0.986$, respectively, and BM refers to CRV-BM CIs. For CRV-BM, BME and BSD, Avg. normalized SE refers to average normalized standard error, described in the text. For EHW, CRV and CRV2, it corresponds to average standard error, averaged over the simulation runs.

Table S5: Simulation results in second Monte Carlo exercise for $\mu(x) = x + .05 \cdot \cos(\pi \cdot x)$.

					Avg. normalized SE							CI coverage rate (%)						
G_+	G_-	N_h	τ_h	$sd(\widehat{\tau})$	EHW	CRV	CRV2	BM	BME	BSD$_{.49}$	BSD$_{.99}$	EHW	CRV	CRV2	BM	BME	BSD$_{.49}$	BSD$_{.99}$
5	5	100	0	0.148	0.149	0.105	0.142	0.222	0.332	0.197	0.249	94.7	78.9	88.1	96.6	100.0	98.4	99.0
		1000		0.046	0.047	0.033	0.044	0.069	0.098	0.091	0.124	95.7	78.4	89.0	96.9	100.0	99.0	99.5
		10000		0.015	0.015	0.012	0.017	0.026	0.033	0.047	0.067	94.8	84.1	92.1	98.2	100.0	99.5	100.0
5	25	100	0	0.143	0.141	0.115	0.136	0.186	0.686	0.178	0.219	94.5	85.4	90.2	96.8	100.0	97.8	97.8
		1000		0.044	0.044	0.036	0.043	0.058	0.248	0.078	0.102	95.5	87.3	91.4	97.0	100.0	98.5	97.8
		10000		0.014	0.014	0.012	0.015	0.020	0.075	0.037	0.052	95.4	88.8	93.2	98.1	100.0	99.1	100.0
5	50	100	0	0.143	0.140	0.115	0.135	0.183	0.754	0.176	0.216	93.7	86.3	90.1	96.1	100.0	97.4	97.9
		1000		0.043	0.044	0.036	0.043	0.057	0.397	0.077	0.100	95.3	87.6	91.6	97.3	100.0	98.9	98.0
		10000		0.014	0.014	0.012	0.015	0.020	0.112	0.036	0.050	95.3	89.4	93.5	97.5	100.0	99.0	99.9
25	25	100	0	0.136	0.131	0.124	0.131	0.143	0.791	0.159	0.186	93.8	91.5	92.9	94.7	100.0	96.6	96.9
		100000		0.040	0.041	0.039	0.041	0.044	0.280	0.064	0.076	95.4	93.1	94.5	95.8	100.0	97.9	98.0
		10000		0.013	0.013	0.013	0.013	0.014	0.084	0.027	0.032	95.4	93.9	95.0	95.9	100.0	97.8	98.3
25	50	100	0	0.132	0.130	0.125	0.130	0.140	0.843	0.156	0.183	94.2	92.9	93.7	95.4	100.0	96.9	96.7
		1000		0.042	0.041	0.039	0.041	0.043	0.430	0.063	0.074	94.7	92.9	93.9	95.1	100.0	97.4	97.7
		10000		0.013	0.013	0.013	0.013	0.014	0.119	0.026	0.031	94.8	93.5	94.1	95.1	100.0	97.9	97.8
50	50	100	0	0.133	0.129	0.126	0.130	0.138	0.880	0.154	0.180	93.6	92.7	93.3	94.6	100.0	96.6	96.7
		1000		0.040	0.041	0.040	0.041	0.042	0.472	0.062	0.072	95.3	94.1	94.8	95.5	100.0	97.6	97.7
		10000		0.013	0.013	0.013	0.013	0.013	0.128	0.025	0.029	94.9	94.1	94.6	95.1	100.0	97.5	97.7

Note: Results are based on 5,000 simulation runs. BSD$_{.49}$ and BSD$_{.99}$ refer to BSD CIs with $K = 0.493$ and $K = 0.986$, respectively, and BM refers to CRV-BM CIs. For CRV-BM, BME and BSD, Avg. normalized SE refers to average normalized standard error, described in the text. For EHW, CRV and CRV2, it corresponds to average standard error, averaged over the simulation runs.

Table S6: Simulation results in second Monte Carlo exercise for $\mu(x) = x + .05 \cdot \sin(\pi \cdot x)$.

					Avg. normalized SE							CI coverage rate (%)						
G_+	G_-	N_h	τ_h	$sd(\hat{\tau})$	EHW	CRV	CRV2	BM	BME	BSD$_{.49}$	BSD$_{.99}$	EHW	CRV	CRV2	BM	BME	BSD$_{.49}$	BSD$_{.99}$
5	5	100	0.108	0.148	0.150	0.107	0.145	0.227	0.334	0.197	0.249	88.6	68.9	81.6	94.0	100.0	96.3	98.5
		1000		0.046	0.047	0.039	0.054	0.084	0.102	0.091	0.124	37.3	27.7	47.5	76.2	96.0	97.2	99.2
		10000		0.015	0.015	0.025	0.036	0.057	0.041	0.047	0.067	0.0	0.1	5.8	54.3	5.9	98.4	99.9
5	25	100	0.090	0.143	0.141	0.116	0.138	0.188	0.687	0.178	0.219	89.4	78.7	84.5	94.2	100.0	96.0	97.5
		1000		0.044	0.044	0.040	0.049	0.065	0.252	0.078	0.102	47.7	39.9	53.0	72.6	100.0	97.0	97.7
		10000		0.014	0.014	0.021	0.028	0.038	0.085	0.037	0.052	0.0	0.2	5.1	25.0	100.0	98.2	99.9
5	50	100	0.088	0.143	0.140	0.116	0.137	0.185	0.755	0.176	0.216	89.3	80.2	85.5	94.0	100.0	95.6	97.6
		1000		0.043	0.044	0.039	0.048	0.064	0.400	0.077	0.100	49.4	41.3	53.9	72.9	100.0	97.1	98.0
		10000		0.014	0.014	0.020	0.028	0.037	0.121	0.036	0.050	0.0	0.3	5.6	25.9	100.0	98.4	99.9
25	25	100	0.072	0.135	0.131	0.125	0.132	0.144	0.792	0.159	0.186	90.5	87.9	89.5	92.1	100.0	95.1	96.4
		100000		0.040	0.041	0.041	0.042	0.045	0.283	0.064	0.076	59.4	57.8	61.4	65.9	100.0	97.0	97.7
		10000		0.013	0.013	0.016	0.017	0.018	0.092	0.027	0.032	0.0	0.2	0.4	0.6	100.0	97.2	98.2
25	50	100	0.070	0.132	0.130	0.125	0.131	0.141	0.844	0.156	0.183	91.0	88.7	90.2	92.6	100.0	95.5	96.4
		1000		0.042	0.041	0.040	0.042	0.044	0.431	0.063	0.074	60.0	58.3	61.1	64.5	100.0	96.3	97.5
		10000		0.013	0.013	0.015	0.016	0.017	0.126	0.026	0.031	0.1	0.1	0.2	0.2	100.0	97.1	97.8
50	50	100	0.068	0.132	0.130	0.126	0.130	0.138	0.881	0.154	0.180	91.0	89.6	90.5	92.2	100.0	95.1	96.4
		1000		0.040	0.041	0.040	0.041	0.043	0.474	0.062	0.072	62.2	61.0	62.8	64.9	100.0	96.4	97.5
		10000		0.013	0.013	0.014	0.015	0.015	0.135	0.025	0.029	0.1	0.2	0.2	0.2	100.0	97.0	97.5

Note: Results are based on 5,000 simulation runs. BSD$_{.49}$ and BSD$_{.99}$ refer to BSD CIs with $K = 0.493$ and $K = 0.986$, respectively, and BM refers to CRV-BM CIs. For CRV-BM, BME and BSD, Avg. normalized SE refers to average normalized standard error, described in the text. For EHW, CRV and CRV2, it corresponds to average standard error, averaged over the simulation runs.

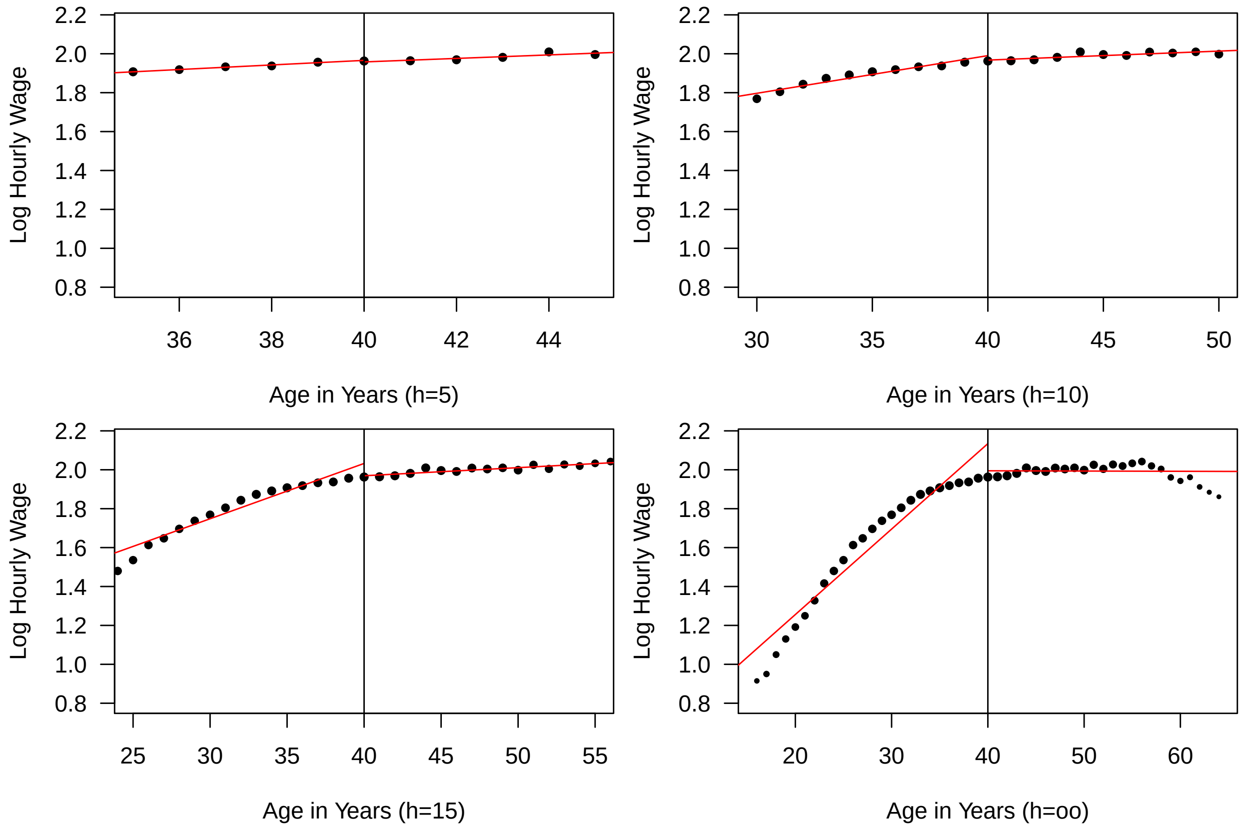

Figure S1: Fit of specification (1) for $p = 1$ (linear, red line) in the full CPS data. The figure displays fit for bandwidths $h = 5$ (top-left panel), $h = 10$ (top-right panel), $h = 15$ (bottom-left panel), and $h = \infty$ (bottom-right panel)

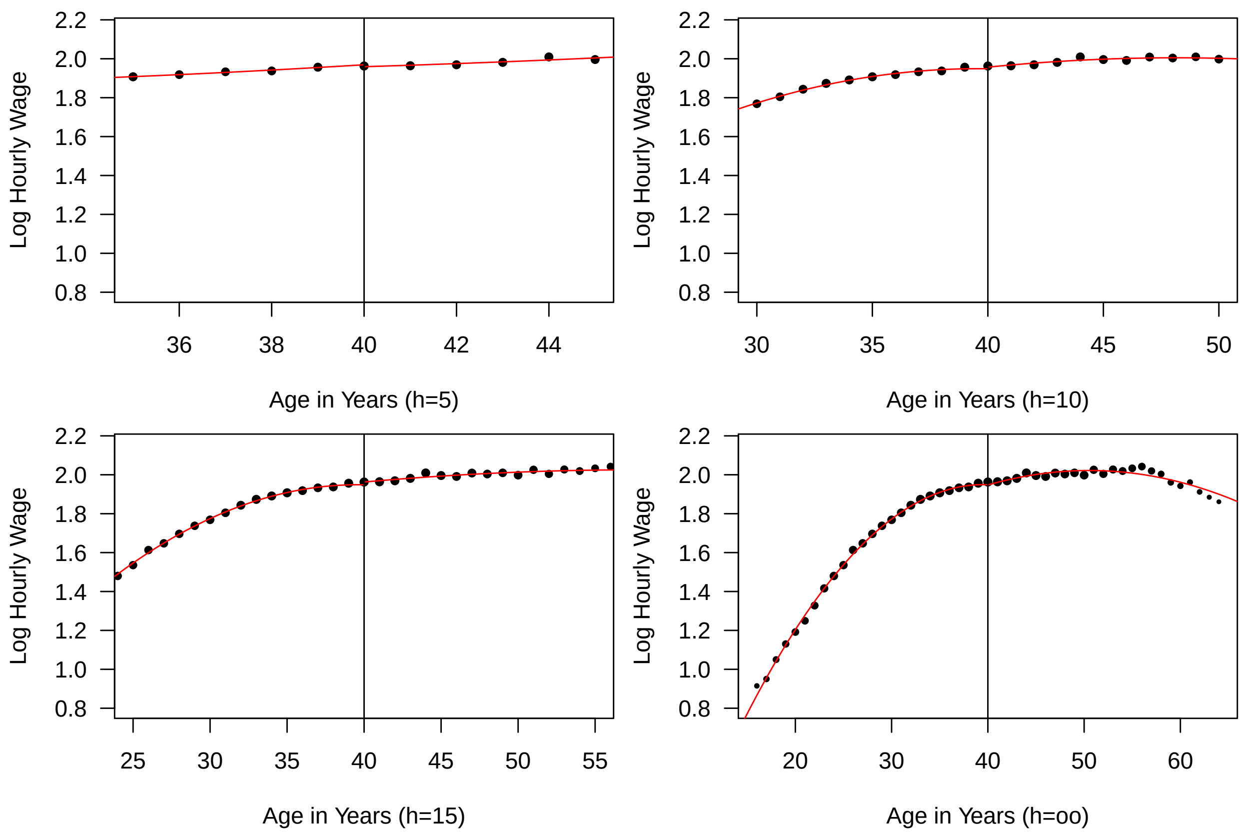

Figure S2: Fit of specification (1) for $p = 2$ (quadratic, red line) in the full CPS data. The figure displays fit for bandwidths $h = 5$ (top-left panel), $h = 10$ (top-right panel), $h = 15$ (bottom-left panel), and $h = \infty$ (bottom-right panel).

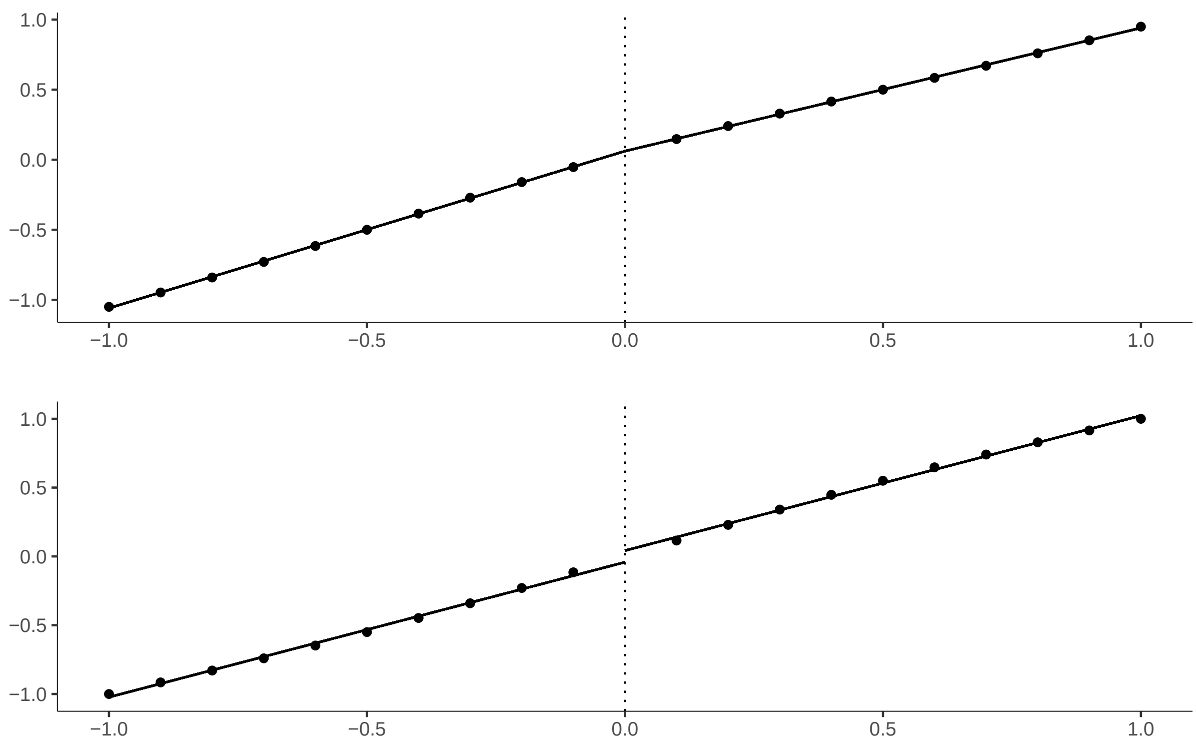

Figure S3: Plot of $\mu(x) = x + .05 \cdot \cos(\pi \cdot x)$ (top panel) and $\mu(x) = x + .05 \cdot \sin(\pi \cdot x)$ (bottom panel) for $G^{-} = G^{+} = 10$. Dots indicate the value of the function at the support points of the running variable; solid lines correspond to linear fit above and below the threshold.