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

Michal Kolesár[*] Christoph Rothe[†]

May 31, 2017

S1. PERFORMANCE OF HONEST CIS IN CPS SIMULATION STUDY

In this section, we again consider the CPS placebo study from Section 2 in the main text to study the performance of the honest CIs proposed in the Section 5. Given that the typical increase in log wages is about 0.017 per extra year of age, guided by the heuristic in Section 5.1, we set $K = 0.045$ for the BSD CIs. 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 to those of EHW and CRV 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.

The coverage properties of honest CIs are excellent: they achieve at least 95% coverage in all specifications. The length of BME CIs is slightly 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 5.2.

S2. EXTRA 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 2.

Each of our DGPs is such that the support of the running variable is the union of an equally spaced grid of G^- points $-1, -(G^- - 1)/G^-, \dots, -1/G^-$ and an equally spaced grid

[*] Woodrow Wilson School and Department of Economics, Princeton University. Electronic correspondence: mkolesar@princeton.edu

[†] Department of Economics, Columbia University. Electronic correspondence: cr2690@columbia.edu

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

$$Y_i = \beta_0 + \tau \cdot \mathbb{I}\{X_i \geq 0\} + \beta^- \cdot X_i + \beta^+ \cdot \mathbb{I}\{X_i \geq 0\} \cdot X_i + U_i. \tag{S1}$$

Note that this specification is analogous to the model (2.1) 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 (S1), we plot the versions of $\mu(x)$ that we consider together with the corresponding linear fit in Figure S1 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 S3–S5 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. 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 $K = \pi^2/10$.[1] To compare length, the tables report normalized standard errors, defined in Section S1.

Table S3 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 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 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 S4 reports results for the case $(\lambda_1, \lambda_2) = (0, .05)$. Here $\mu(x)$ is nonlinear, but due

[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.

to the symmetry properties of the cosine function $\tau_h = 0$. This setup mimics applications in which the bias of $\hat{\tau}$ 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 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 S5 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 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 3 and 4 in the main text. 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 5.2, they are best suited to applications with a small number of 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.

S3. 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 5.1 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 S1. *Suppose $\mu \in \mathcal{M}_H(K)$ for some $K \geq 0$. Then $K \geq |\Delta(x_1, x_2, x_3)|$, where*

$$\Delta(x_1, x_2, x_3) = 2\frac{(1-\lambda)\mu(x_3) + \lambda\mu(x_1) - \mu(x_2)}{(1-\lambda)x_3^2 + \lambda x_1^2 - x_2^2} = 2\frac{(1-\lambda)\mu(x_3) + \lambda\mu(x_1) - \mu(x_2)}{(1-\lambda)\lambda(x_3 - x_1)^2}. \quad \text{(S2)}$$

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

$$\mu(x) = \mu(0) + \mu'(0)x + r(x) \qquad r(x) = \int_0^x (x-u)\mu''(u)\,\mathrm{d}u. \quad \text{(S3)}$$

This follows by an argument given in the proof of Proposition 1 in Appendix B. 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 S1 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 5.1.

While it is possible to estimate a lower bound on K using Lemma S1 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 (S2) 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 small number of 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 S1 will be useful.

Lemma S2. *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*

$$\Delta(I_1, I_2, I_3) = 2\frac{\lambda_n E_n\mu(x_1) + (1-\lambda_n)E_n\mu(x_3) - E_n\mu(x_2)}{E_n\left[(1-\lambda_n)x_3^2 + \lambda_n x_1^2 - x_2^2\right]}. \quad \text{(S4)}$$

Proof. By Equation (S3), 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

$$\begin{aligned}&\lambda_n E_n\mu(x_1) + (1-\lambda_n)E_n\mu(x_3) - E_n\mu(x_2) \\ &\qquad = (1-\lambda_n)E_n\int_{\bar{x}_2}^{x_3}(x_3-u)\mu''(u)\,\mathrm{d}u + \lambda_n E_n\int_{x_1}^{\underline{x}_2}(u-x_1)\mu''(u)\,\mathrm{d}u + S, \quad \text{(S5)}\end{aligned}$$

where

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

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 $q'(u) = \lambda_n - E_n\mathbb{I}\left\{x_2 \leq u\right\}$, so that the first derivative is positive at first and changes sign only once. Therefore, the right-hand side of (S5) 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 (S4) 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 $\bar{y}(I) = \sum_{i=1}^N Y_i\mathbb{I}\left\{X_i \in I\right\}/n(I)$, $n(I) = \sum_{i=1}^N \mathbb{I}\left\{X_i \in I\right\}$, and $\bar{x}^2(I) = \sum_{i=1}^N X_i\mathbb{I}\left\{X_i \in I\right\}/n(I)$, 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 S2 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 $\hat{\Delta}_s = (\hat{\Delta}^-_{1s}, \dots, \hat{\Delta}^-_{M_-s}, \hat{\Delta}^+_{1s}, \dots, \hat{\Delta}^+_{M_+s})$, 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 6 of the main text.

REFERENCES

Armstrong, T. B. and M. Kolesár (2017): "Optimal inference in a class of regression models," ArXiv:1511.06028.

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

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

h	τ_h	N_h	$SD(\hat{\tau})$	Avg. normalized SE				Rate CRV SE	CI coverage rate			
				CRV	EHW	BME	BSD	>EHW SE	CRV	EHW	BME	BSD
5	-0.008	100	0.238	0.167	0.234	0.609	0.309	0.144	0.776	0.943	$>.999$	0.973
		500	0.104	0.073	0.104	0.253	0.170	0.134	0.780	0.947	$>.999$	0.979
		2000	0.052	0.036	0.052	0.125	0.109	0.126	0.773	0.949	$>.999$	0.997
		10000	0.021	0.015	0.023	0.055	0.064	0.093	0.772	0.961	$>.999$	0.999
10	-0.023	100	0.227	0.193	0.223	0.946	0.394	0.256	0.873	0.939	$>.999$	0.970
		500	0.099	0.086	0.099	0.385	0.215	0.249	0.876	0.944	$>.999$	0.980
		2000	0.049	0.044	0.050	0.187	0.131	0.268	0.860	0.930	$>.999$	0.991
		10000	0.021	0.021	0.022	0.086	0.077	0.394	0.781	0.829	$>.999$	0.991
15	-0.063	100	0.222	0.197	0.216	1.137	0.453	0.305	0.884	0.927	$>.999$	0.968
		500	0.095	0.089	0.096	0.507	0.242	0.336	0.853	0.899	$>.999$	0.985
		2000	0.048	0.047	0.048	0.243	0.146	0.453	0.712	0.730	$>.999$	0.987
		10000	0.020	0.028	0.021	0.118	0.087	0.917	0.348	0.153	$>.999$	0.987
∞	-0.140	100	0.212	0.196	0.209	1.236	0.492	0.348	0.874	0.909	$>.999$	0.962
		500	0.093	0.092	0.093	0.664	0.261	0.460	0.746	0.766	$>.999$	0.980
		2000	0.046	0.054	0.046	0.328	0.157	0.816	0.434	0.306	$>.999$	0.984
		10000	0.020	0.038	0.021	0.184	0.094	1.000	0.036	0.000	$>.999$	0.982

Note: Results are based on 10,000 simulation runs. For BSD, $M = 0.045$. For BME and BSD, Avg. normalized SE refers to average normalized standard error, described in the text. For EHW and CRV, 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				Rate CRV SE	CI coverage rate			
h	N_h	τ_h	$sd(\hat{\tau})$	CRV	EHW	BME	BSD	>EHW SE	CRV	EHW	BME	BSD
				Quadratic Specification $(p = 2)$								
5	100	-0.010	0.438	0.206	0.427	0.705	0.309	0.028	0.607	0.932	0.995	0.973
	500		0.189	0.086	0.190	0.301	0.170	0.012	0.599	0.947	0.997	0.979
	2000		0.093	0.042	0.095	0.149	0.109	0.007	0.587	0.951	0.998	0.997
	10000		0.038	0.018	0.042	0.067	0.064	0.004	0.595	0.964	0.999	0.999
10	100	0.008	0.361	0.258	0.349	1.000	0.394	0.147	0.795	0.933	$> .999$	0.970
	500		0.157	0.110	0.156	0.405	0.215	0.116	0.790	0.948	$> .999$	0.980
	2000		0.077	0.055	0.078	0.196	0.131	0.112	0.794	0.947	$> .999$	0.991
	10000		0.033	0.025	0.035	0.087	0.077	0.126	0.808	0.956	$> .999$	0.991
15	100	0.014	0.349	0.270	0.329	1.189	0.453	0.205	0.836	0.923	$> .999$	0.968
	500		0.146	0.117	0.147	0.516	0.242	0.179	0.851	0.946	$> .999$	0.985
	2000		0.073	0.058	0.073	0.243	0.146	0.168	0.839	0.946	$> .999$	0.987
	10000		0.031	0.026	0.033	0.107	0.087	0.175	0.828	0.937	$> .999$	0.987
∞	100	-0.001	0.327	0.269	0.313	1.286	0.492	0.245	0.862	0.926	$> .999$	0.962
	500		0.141	0.119	0.140	0.654	0.261	0.221	0.865	0.943	$> .999$	0.980
	2000		0.069	0.059	0.070	0.300	0.157	0.215	0.877	0.948	$> .999$	0.984
	10000		0.030	0.027	0.031	0.135	0.094	0.239	0.861	0.936	$> .999$	0.982

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

Table S3: Simulation results in second Monte Carlo exercise for $\mu(x) = x$.

					Avg. normalized SE					Rate CRV	CI coverage rate				
G_+	G_-	N_h	τ_h	$sd(\hat{\tau})$	CRV	EHW	BME	BSD$_{.49}$	BSD$_{.98}$	>EHW SE	CRV	EHW	BME	BSD$_{.49}$	BSD$_{.98}$
5	5	100	0	0.148	0.105	0.149	0.332	0.197	0.249	0.142	0.791	0.946	$> .999$	0.985	0.990
		1000		0.046	0.032	0.047	0.097	0.091	0.124	0.116	0.775	0.957	$> .999$	0.990	0.995
		10000		0.015	0.010	0.015	0.031	0.047	0.067	0.122	0.773	0.948	$> .999$	0.995	1.000
5	25	100	0	0.143	0.114	0.141	0.686	0.178	0.219	0.165	0.854	0.945	$> .999$	0.978	0.978
		1000		0.044	0.036	0.044	0.248	0.078	0.102	0.154	0.872	0.955	$> .999$	0.986	0.981
		10000		0.014	0.011	0.014	0.074	0.037	0.052	0.151	0.867	0.955	$> .999$	0.994	1.000
5	50	100	0	0.143	0.115	0.140	0.754	0.176	0.216	0.150	0.860	0.937	$> .999$	0.974	0.978
		1000		0.043	0.036	0.044	0.397	0.077	0.100	0.134	0.873	0.953	$> .999$	0.990	0.981
		10000		0.014	0.011	0.014	0.111	0.036	0.050	0.140	0.873	0.952	$> .999$	0.992	0.999
25	25	100	0	0.136	0.124	0.131	0.791	0.159	0.186	0.342	0.913	0.938	$> .999$	0.966	0.969
		1000		0.040	0.039	0.041	0.280	0.064	0.076	0.344	0.932	0.955	$> .999$	0.979	0.980
		10000		0.013	0.012	0.013	0.083	0.027	0.032	0.338	0.931	0.954	$> .999$	0.978	0.983
25	50	100	0	0.132	0.125	0.130	0.843	0.156	0.183	0.361	0.928	0.942	$> .999$	0.969	0.967
		1000		0.042	0.039	0.041	0.430	0.063	0.074	0.354	0.930	0.947	$> .999$	0.974	0.976
		10000		0.013	0.012	0.013	0.119	0.026	0.031	0.340	0.932	0.949	$> .999$	0.979	0.978
50	50	100	0	0.133	0.126	0.129	0.879	0.154	0.180	0.394	0.926	0.935	$> .999$	0.966	0.967
		1000		0.040	0.040	0.041	0.472	0.062	0.072	0.390	0.941	0.953	$> .999$	0.976	0.977
		10000		0.013	0.012	0.013	0.127	0.025	0.029	0.374	0.939	0.949	$> .999$	0.975	0.977

Note: Results are based on 5,000 simulation runs. BSD$_{.49}$ and BSD$_{.98}$ refer to BSD CIs with $K = 0.49$ and $K = 0.98$, respectively. For BME and BSD, Avg. normalized SE refers to average normalized standard error, described in the text. For EHW and CRV, it corresponds to average standard error, averaged over the simulation runs.

Table S4: Simulation results in second Monte Carlo exercise for $\mu(x) = x + .05 \cdot \cos(\pi \cdot x)$.

					Avg. normalized SE					Rate CRV	CI coverage rate				
G_+	G_-	N_h	τ_h	$sd(\hat{\tau})$	CRV	EHW	BME	$BSD_{.49}$	$BSD_{.98}$	>EHW SE	CRV	EHW	BME	$BSD_{.49}$	$BSD_{.98}$
5	5	100	0	0.148	0.105	0.149	0.332	0.197	0.249	0.143	0.789	0.947	$>.999$	0.984	0.990
		1000		0.046	0.033	0.047	0.098	0.091	0.124	0.126	0.784	0.957	$>.999$	0.990	0.995
		10000		0.015	0.012	0.015	0.033	0.047	0.067	0.211	0.841	0.948	$>.999$	0.995	1.000
5	25	100	0	0.143	0.115	0.141	0.686	0.178	0.219	0.166	0.854	0.945	$>.999$	0.978	0.978
		1000		0.044	0.036	0.044	0.248	0.078	0.102	0.163	0.873	0.955	$>.999$	0.985	0.978
		10000		0.014	0.012	0.014	0.075	0.037	0.052	0.239	0.888	0.954	$>.999$	0.991	1.000
5	50	100	0	0.143	0.115	0.140	0.754	0.176	0.216	0.150	0.863	0.937	$>.999$	0.974	0.979
		1000		0.043	0.036	0.044	0.397	0.077	0.100	0.143	0.876	0.953	$>.999$	0.989	0.980
		10000		0.014	0.012	0.014	0.112	0.036	0.050	0.209	0.894	0.953	$>.999$	0.990	0.999
25	25	100	0	0.136	0.124	0.131	0.791	0.159	0.186	0.344	0.915	0.938	$>.999$	0.966	0.969
		1000		0.040	0.039	0.041	0.280	0.064	0.076	0.347	0.931	0.954	$>.999$	0.979	0.980
		10000		0.013	0.013	0.013	0.084	0.027	0.032	0.398	0.939	0.954	$>.999$	0.978	0.983
25	50	100	0	0.132	0.125	0.130	0.843	0.156	0.183	0.361	0.929	0.942	$>.999$	0.969	0.967
		1000		0.042	0.039	0.041	0.430	0.063	0.074	0.354	0.929	0.947	$>.999$	0.974	0.977
		10000		0.013	0.013	0.013	0.119	0.026	0.031	0.383	0.935	0.948	$>.999$	0.979	0.978
50	50	100	0	0.133	0.126	0.129	0.880	0.154	0.180	0.394	0.927	0.936	$>.999$	0.966	0.967
		1000		0.040	0.040	0.041	0.472	0.062	0.072	0.394	0.941	0.953	$>.999$	0.976	0.977
		10000		0.013	0.013	0.013	0.128	0.025	0.029	0.420	0.941	0.949	$>.999$	0.975	0.977

Note: Results are based on 5,000 simulation runs. $BSD_{.49}$ and $BSD_{.98}$ refer to BSD CIs with $K = 0.49$ and $K = 0.98$, respectively. For BME and BSD, Avg. normalized SE refers to average normalized standard error, described in the text. For EHW and CRV, it corresponds to average standard error, averaged over the simulation runs.

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

					Avg. normalized SE					Rate CRV	CI coverage rate				
G_+	G_-	N_h	τ_h	$sd(\hat{\tau})$	CRV	EHW	BME	BSD$_{.49}$	BSD$_{.98}$	>EHW SE	CRV	EHW	BME	BSD$_{.49}$	BSD$_{.98}$
5	5	100	0.11	0.148	0.107	0.150	0.334	0.197	0.249	0.157	0.689	0.886	$> .999$	0.963	0.985
		1000		0.046	0.039	0.047	0.102	0.091	0.124	0.259	0.277	0.373	0.960	0.972	0.992
		10000		0.015	0.025	0.015	0.041	0.047	0.067	0.962	0.001	0.000	0.059	0.984	0.999
5	25	100	0.09	0.143	0.116	0.141	0.687	0.178	0.219	0.171	0.787	0.894	$> .999$	0.960	0.975
		1000		0.044	0.040	0.044	0.252	0.078	0.102	0.278	0.399	0.477	$> .999$	0.970	0.977
		10000		0.014	0.021	0.014	0.085	0.037	0.052	0.931	0.002	0.000	$> .999$	0.982	0.999
5	50	100	0.09	0.143	0.116	0.140	0.755	0.176	0.216	0.159	0.802	0.893	$> .999$	0.956	0.976
		1000		0.043	0.039	0.044	0.400	0.077	0.100	0.252	0.413	0.494	$> .999$	0.971	0.980
		10000		0.014	0.020	0.014	0.121	0.036	0.050	0.906	0.003	0.000	$> .999$	0.984	0.999
25	25	100	0.07	0.135	0.125	0.131	0.792	0.159	0.186	0.349	0.879	0.905	$> .999$	0.951	0.964
		1000		0.040	0.041	0.041	0.283	0.064	0.076	0.432	0.578	0.594	$> .999$	0.970	0.977
		10000		0.013	0.016	0.013	0.092	0.027	0.032	0.933	0.002	0.000	$> .999$	0.972	0.982
25	50	100	0.07	0.132	0.125	0.130	0.844	0.156	0.183	0.370	0.887	0.910	$> .999$	0.955	0.964
		1000		0.042	0.040	0.041	0.431	0.063	0.074	0.429	0.583	0.600	$> .999$	0.963	0.975
		10000		0.013	0.015	0.013	0.126	0.026	0.031	0.901	0.001	0.001	$> .999$	0.971	0.978
50	50	100	0.07	0.132	0.126	0.130	0.881	0.154	0.180	0.401	0.896	0.910	$> .999$	0.951	0.964
		1000		0.040	0.040	0.041	0.474	0.062	0.072	0.458	0.610	0.622	$> .999$	0.964	0.975
		10000		0.013	0.014	0.013	0.135	0.025	0.029	0.868	0.002	0.001	$> .999$	0.970	0.975

Note: Results are based on 5,000 simulation runs. BSD$_{.49}$ and BSD$_{.98}$ refer to BSD CIs with $K = 0.49$ and $K = 0.98$, respectively. For BME and BSD, Avg. normalized SE refers to average normalized standard error, described in the text. For EHW and CRV, it corresponds to average standard error, averaged over the simulation runs.

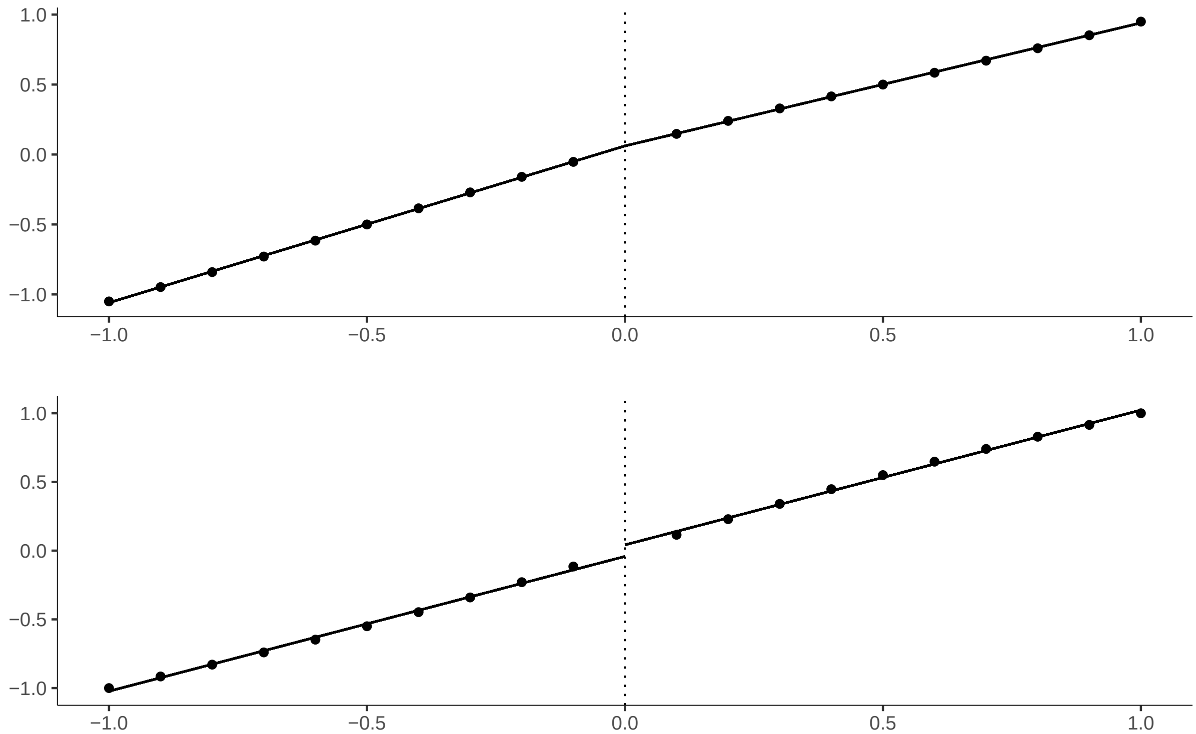

Figure S1: Plot of $\mu(x) = x + .05 \cdot \cos(\pi \cdot x)$ (top panel) and $\mu(x) = x + .05 \cdot \sin(\pi \cdot x)$ (bottom panel) for $G^- = G^+ = 10$. Dots indicate the value of the function at the support points of the running variable; solid lines correspond to linear fit above and below the threshold.