Online Supplement: Flexible Covariate Adjustments in Regression Discontinuity Designs

Claudia Noack Tomasz Olma Christoph Rothe

Abstract

This Online Supplement contains additional empirical and simulation results.

S1. ADDITIONAL EMPIRICAL RESULTS

In Figure S1, we present the full results of our empirical analysis for bias-aware inference and robust bias correction. The first two graphs of Panel A are discussed in the main text. The third and fourth graphs illustrate the length of the confidence intervals associated with the cross-fitted and conventional linear covariate adjustments, respectively, relative to the no covariates confidence intervals. We note that the conventional linear adjustment yields on average slightly shorter confidence intervals, but this effect might be due the downward bias of the associated standard error documented in Simulation II in Section 7. Panel B presents the results based on the bias-aware approach with the second-stage smoothness bound calibrated based on the adjusted outcomes using the rule of thumb of Imbens and Wager (2019). This choice was dictated by practical considerations, as it would not be possible to separately discuss the choice of smoothness bound for each of the 56 specifications. In comparison to Panel A, all four histograms are more spread out, which reflects the differences between the smoothness bounds calibrated based on the original and adjusted outcomes. In some cases, the confidence intervals based on the flexible adjustment are wider than the no covariates and linear adjustment confidence intervals, which is due to an increase in the smoothness bound, but the average reductions in the confidence interval length are larger then in Panel A. We note, however, that these comparisons are sensitive to the method of choosing the smoothness bound. Panel C

First version: July 16, 2021. This version: December 2, 2024. The authors gratefully acknowledge financial support by the European Research Council (ERC) through grant SH1-77202. The second author also gratefully acknowledges support from the European Research Council through Starting Grant No. 852332. Author contact information: Claudia Noack, Department of Economics, University of Bonn. E-Mail: claudia.noack@uni-bonn.de. Website: https://claudianoack.github.io. Tomasz Olma, Department of Statistics, Ludwig Maximilian University of Munich. E-Mail: t.olma@lmu.de. Website: https://tomaszolma.github.io. Christoph Rothe, Department of Economics, University of Mannheim. E-Mail: rothe@vwl.uni-mannheim.de. Website: http://www.christophrothe.net.

presents the results based on the robust bias correction. They are qualitatively similar to those in Panel B. In particular, the flexible and linear adjustments lead to wider confidence intervals in some cases. By inspecting the results, we saw that these increases occurred in cases where the adjusted bandwidth was smaller than the no covariates bandwidths.

In 13 out of 16 papers in our literature analysis, the standard errors were clustered. To account for that, in Figure S2, we present the results of our empirical analysis with clustered standard errors. In our second-stage RD regression, we cluster the standard error based on the same variable as in the original application. Additionally, we adjust the data splitting procedure such that all observations within a cluster belong to the same fold. Clustering substantially increases the length of all confidence intervals, but the relative patterns displayed in Figure S2 are broadly similar to those in Figure S1.

S2. ADDITIONAL SIMULATION RESULTS

In this section, we provide more details and additional results for the simulation studies in Section 7.

S2.1. **Scope for Efficiency Gains.** To gauge the scope for efficiency gains due to covariate adjustments in this simulation setting, in Table S2, we present RD estimates at the placebo cutoff using all the observations in the restricted data set of Londoño-Vélez et al. (2020) described in Section 7.1. As in in the main text, we consider the original outcome and age as the dependent variables. We now employ the robust bias correction approach in addition to the bias-aware inference. In Panel A, the results are very similar in all rows, which indicates that the covariates have virtually no explanatory power for the outcome and so the covariate adjustments do not lead to meaningful changes in the length of confidence intervals. When considering the age as the dependent variable in Panel B, the machine learning adjustments improve upon the no covariates and linear adjustment RD estimators, with our proposed flexible adjustment leading to the shortest confidence intervals.

S2.2. **Additional Results for Simulation I.** Table S3 extends the results in Table 1 from the main text and displays the results for all individual methods considered in our flexible adjustment. For all methods that employ cross-fitting, we consider their oracle versions obtained on the restricted data set. The observations about the performance of the flexible adjustment discussed in the main text apply here too. The confidence intervals are slightly conservative, the average standard error is very close to the standard deviation in all cases, and the changes in the bias across different adjustments are minimal relative to the standard deviation. The feasible and infeasible, oracle versions of the estimators perform very similarly. The flexible adjustment consistently leads to

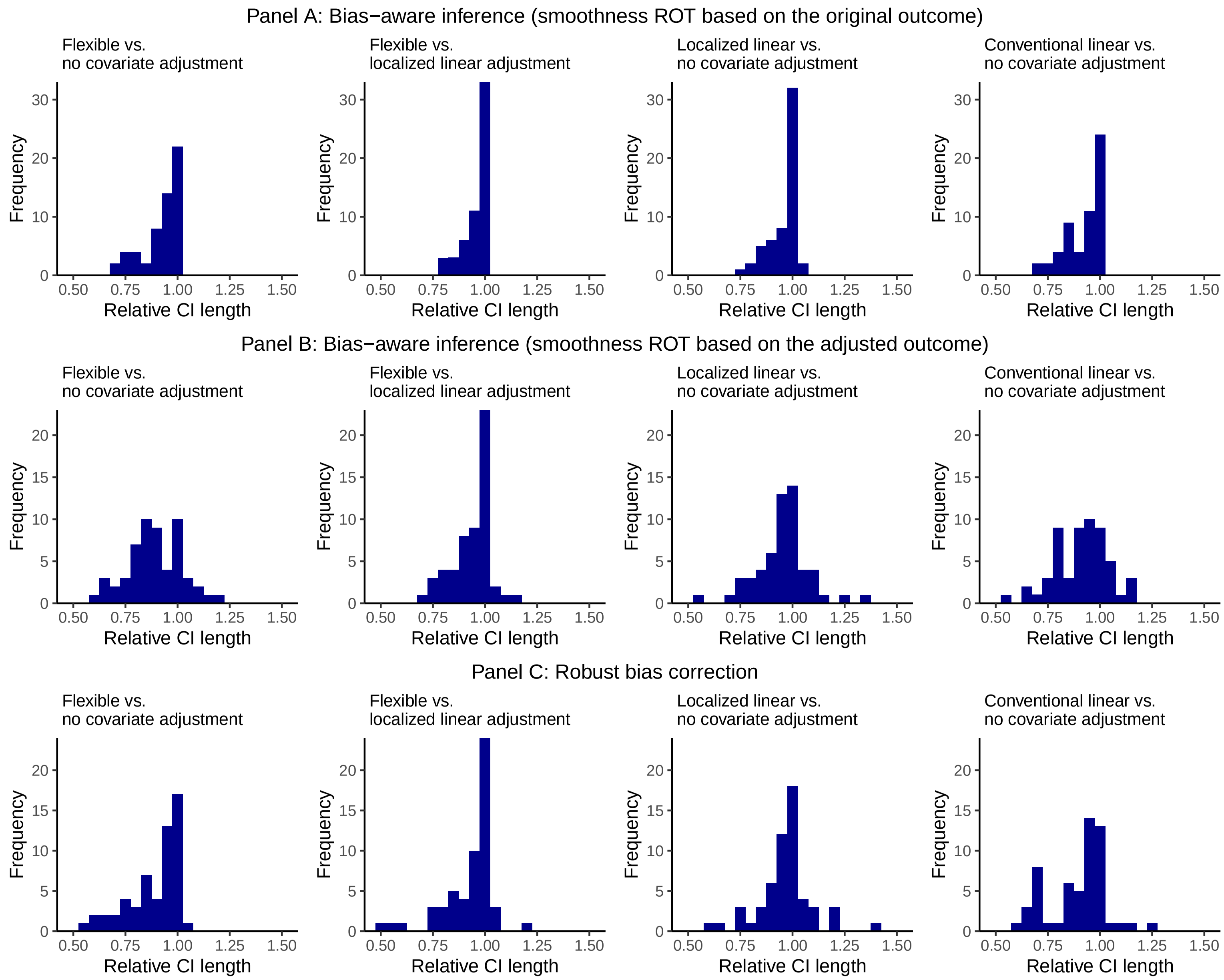

Figure S1: Full empirical results without clustering.

Notes: Results of our empirical analysis for bias-aware inference (in two variants) and robust bias correction. See Section 3 and Appendix B for details on the estimators and the confidence intervals; and Section 6 and Appendix C for details on the data sets.

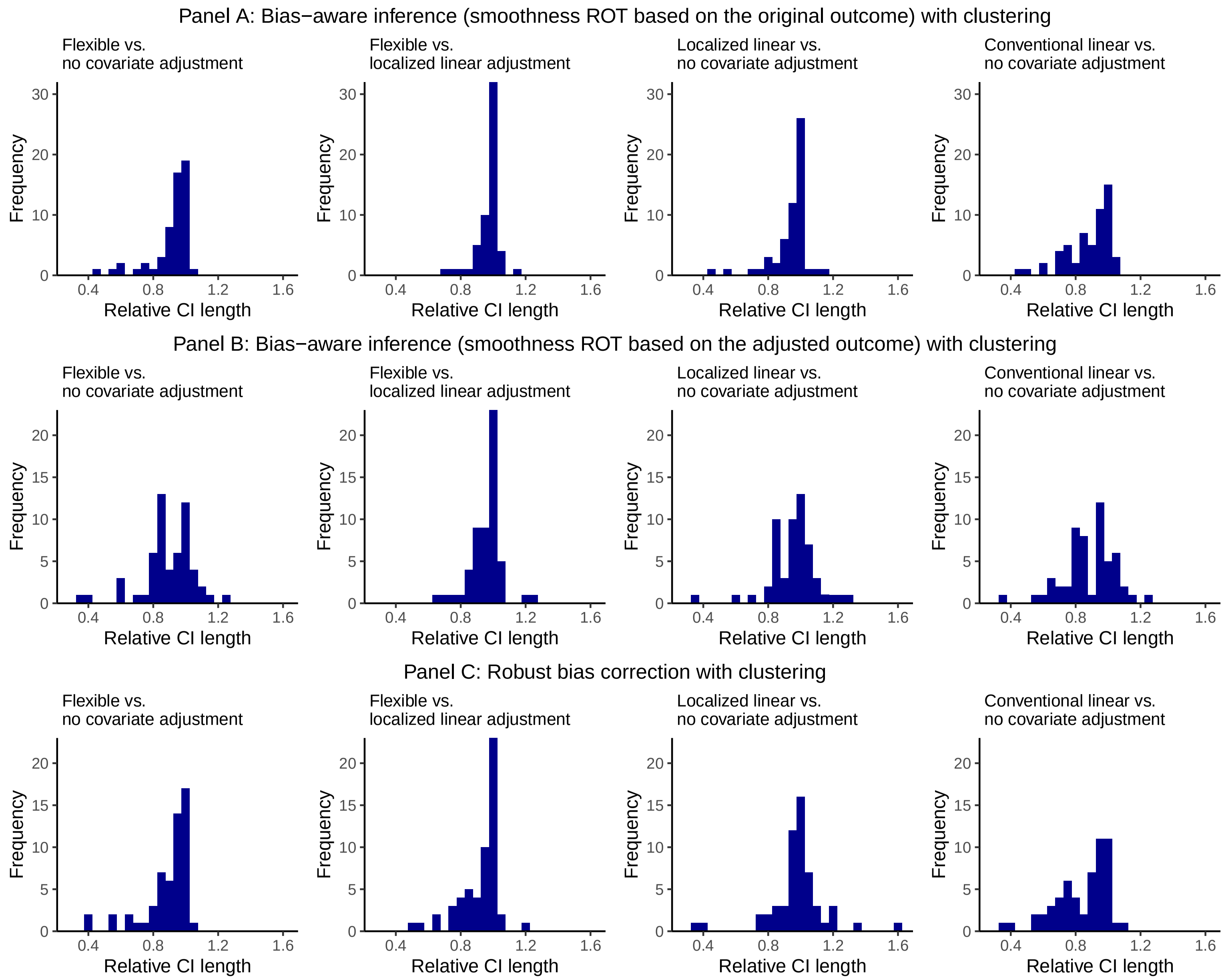

Figure S2: Full empirical results with clustering.

Notes: Results of our empirical analysis for bias-aware inference (in two variants) and robust bias correction with clusered standard errors. See Section 3 and Appendix B for details on the estimators and the confidence intervals; and Section 6 and Appendix C for details on the data sets.

Table S1: Overview of the papers of the literature analysis.

	Outcome	Running variable	Covariates	#Covs (not 0/1)	#Obs	Eff #Obs	Eff #Obs /#Covs	#Clusters	Covs in RD
Akhtari et al. (AER 2022)									
Table 3: "Political Turnover and Fourth Grade and Eighth-grade Test Scores"									
1	Fourth-grade test scores	Incumbent's vote margin	Baseline school-level average test scores; school- and individual-level controls; election-cycle indicator	14 (1)	1,088,553	325,554	23,254	3737	Yes
2	Eighth grade test scores			14 (1)	446,451	234,629	17,545	2368	Yes
Altindag et al. (AEJAE 2022)									
Table 4: "Effects of Curfew on Mental Health Outcomes"									
3	Mental distress	number of months older than index month	month, province, and surveyor fixed effects, indicators for education levels, ethnicity, and gender	175 (0)	1868	475	2.7	144	Yes
4	Somatic symptoms of distress			175 (0)	1868	503	2.8	144	Yes
5	Nonsomatic symptoms of distress			175 (0)	1868	478	2.7	144	Yes
6	Sum of Yes answers in SRQ-20			175 (0)	1868	475	2.7	144	Yes

Notes: The authors present results for different bandwidths. Here, the reported effective sample sizes correspond to the bandwidth calculated via the algorithm of Calonico et al. (2014). In the main specification, month fixed effects are included and the standard error is clustered on the running variable. In our reanalysis, we do not include covariates that are a deterministic function of the running variable

	Outcome	Running variable	Covariates	#Covs (not 0/1)	#Obs	Eff #Obs	Eff #Obs /#Covs	#Clusters	Covs in RD
Ambrus et al. (AER 2020)									
Table 3: "Boundary Effects of Rental Prices"									
7	Log rental prices, 1853	Distance to boundary	Determinants of rental values, distance to various amenities, distance to presumed plague pit, and sewer access	14 (12)	1738	469	34	179	Yes
8	Log rental prices, 1864			14 (12)	1738	510	36	179	Yes
9	Log rental prices, 1894			5 (5)	1879	363	73	179	Yes
10	Log rental prices, 1936			6 (6)	793	221	37	90	Yes

Table S1: Overview of the papers of the literature analysis.

	Outcome	Running variable	Covariates	#Covs (not 0/1)	#Obs	Eff #Obs	Eff #Obs /#Covs	#Clusters	Covs in RD
Asher and Novosad (AER 2020)									
Table 3: “Impact of New Road on Indices of Major Outcomes”									
11	Transportation	Village population	baseline village-level controls for amenities and economic indicators, as well as district-cutoff fixed effects	225 (8)	11432	11432	51	-	Yes
12	Occupation			225 (8)	11432	11432	51	-	Yes
13	Firms			225 (8)	10678	10678	48	-	Yes
14	Production			225 (8)	11432	11432	51	-	Yes
15	Consumption			225 (8)	11432	11432	51	-	Yes
Avis et al. (AEJAE 2022)									
Table 4: “Effects of Campaign Spending Limits on Candidate Entry”									
16	# of candidates	maximum amount a candidate spent in municipality election	Municipal controls: GDP per capita, illiteracy, share urban, Gini coefficient, population	5 (5)	5562	3080	616	-	Yes
17	Eff. # of candidates			5 (5)	5558	3052	610	-	Yes
18	Small party			5 (5)	5562	3116	623	-	Yes
19	Small party w/o incumbent			5 (5)	5562	2804	561	-	Yes
20	Party’s ideology index			5 (5)	5562	2783	557	-	Yes
21	Candidate’s prop. to win			5 (5)	5459	3074	615	-	Yes
22	Candidate’s wealth			5 (5)	5562	3218	644	-	Yes
23	Candidate’s political experience			5 (5)	5562	2849	570	-	Yes
24	Candidate’s gender			5 (5)	5562	3080	616	-	Yes
25	Candidate’s age			5 (5)	5562	3259	652	-	Yes
26	Candidate’s college degree			5 (5)	5562	2881	576	-	Yes
27	Candidate: white			5 (5)	5562	2668	534	-	Yes

Table S1: Overview of the papers of the literature analysis.

	Outcome	Running variable	Covariates	#Covs (not 0/1)	#Obs	Eff #Obs	Eff #Obs /#Covs	#Clusters	Covs in RD
Baskaran and Hessami (AEJEP 2018)									
Table 2: "Baseline Results: Rank Improvement of Female Candidates"									
28	Rank improvement	vote margin	municipality characteristics	-	6472	2878	-	134	No
Notes: We use 24 (24 non-binary) covariates from the robustness check in Table A.4.									
Becker et al. (AER 2020)									
Table A.10: "Border Sample from the Diagnoza Survey"									
29	Years of education	distance to boarder	Respondents gender, age, squared age, dummies for six age groups, indicators for Western Territories, rural places and urban counties	20 (17)	33160	8760	438	11734	Yes
Notes: All RD results are in the appendix.									

Table S1: Overview of the papers of the literature analysis.

	Outcome	Running variable	Covariates	#Covs (not 0/1)	#Obs	Eff #Obs	Eff #Obs /#Covs	#Clusters	Covs in RD
Chin (AEJAE 2023)									
Table 2: “Effect on the Geographic Concentration of Voters ”and Table 4 Panel C in the Appendix									
30	Coefficient of variation of voters	Number of registered voters	Election-year fixed effects, municipality characteristics	17 (11)	22915	230	14	5568	Yes
31	Fractionalization of voters			17 (11)	33187	230	14	5568	Yes
32	Entropy of voters			17 (11)	33187	230	14	5568	Yes
33	SD in vote shares for first placed			17 (11)	33187	230	14	5568	Yes
34	SD in vote shares for second placed			17 (11)	33187	230	14	5568	Yes
35	SD in vote shares for third placed			17 (11)	33187	217	13	5568	Yes
36	SD in vote shares for fourth placed			17 (11)	33187	185	11	5568	Yes
Notes: Additionally to the covariates used in the main text, we use all covariates that were used in Table 4 Panel C in the Appendix. As the number of observations of the original data set is very large and its distribution is very skewed around the cutoff, we restricted the sample to lie within three times of the bandwidth used in the main analysis around the cutoff. In the main specification, the author include the density of the population as a control, but we don’t do this.									
Curto-Grau et al. (AEJAE 2018)									
Table 1 A: “Average Effect of Partisan Alignment on Capital Transfers”									
37	Alignment	incumbent’s vote margin	Financial and demographic municipality characteristics,	14 (0)	6050	2553	102	2592	Yes
Notes: In their main specification, they include 14 fixed effects. We do not use them in our no-covariates RD estimator. For our RD estimators that use covariates, we also include all covariates that are used for from the falsification check of Figure A.10. This gives us a total of 25 (10 non-binary) covariates.									

Table S1: Overview of the papers of the literature analysis.

	Outcome	Running variable	Covariates	#Covs (not 0/1)	#Obs	Eff #Obs	Eff #Obs /#Covs	#Clusters	Covs in RD
Granzier et al. (AEJAE 2023)									
Table 2: “Impact on Running in the Second Round and Winning”and Table C4									
38	Running	Vote Margin 1 vs 2	gender, characteristics of previous election and party, incumbent, strength	23 (8)	45064	24544	1067	8970	OA
39	Winning	Vote Margin 1 vs 2		23 (8)	45064	16054	698	8970	OA
40	Running	Vote Margin 2 vs 3		23 (8)	17730	10694	465	4810	OA
41	Winning	Vote Margin 2 vs 3		23 (8)	17730	8796	382	4810	OA
42	Running	Vote Margin 3 vs 4		23 (8)	3956	2338	102	1243	OA
43	Winning	Vote Margin 3 vs 4		23 (8)	3956	2232	97	1243	OA
Greenstone et al. (AER Insigts 2022)									
Table 1: “Automating Air Quality Monitoring System and Reported PM_{10}”, Column 2									
44	PM_{10} concentration	Days to automation	weather controls, and station and month fixed effects	670 (4)	1,049,325	49,843	74	123	Yes
Notes: We do not include covariates that are determined based on the running variable and therefore exclude month fixed effects from our analysis.									
Johnson (AER 2020)									
Table 2: “Instrumental Variables (IV) Estimate of the General Deterrence Effect of a Press Release on Compliance of Other Facilities ”and Table A.1									
45	Number of Violations	Focal penalty		2 (0)	60,416	3302	1651	2746	Yes
46	Number of Violations	Focal penalty	construction, programmed	1 (0)	39,058	10,873	10,873	2455	Yes

Notes: Second specification excludes inspections initiated by a serious accident worker complaint, or referral.
We further consider 30 covariates that are from Table A.1 (press release, cfr, union, # inspection prior tc, total violations prior tc).

Table S1: Overview of the papers of the literature analysis.

Tuttle (AEJEP 2019)

Table 3: “Main Results: Effect of the SNAP Ban on Recidivism”

	Outcome	Running variable	Covariates	#Covs (not 0/1)	#Obs	Eff #Obs	Eff #Obs /#Covs	#Clusters	Covs in RD
47	Recidivism			14 (4)	18850	790	56	5385	OA
48	Financially motivated recidivism	Date	offender characteristics	14 (4)	18850	936	67	5385	OA
49	Non-financially motivated recidivism			14 (4)	18850	980	70	5385	OA

Notes: In the main specification, dummies for weekdays are included and the standard error is clustered on the running variable. We do not cluster the standard errors on the running variable and we do not include covariates that are a deterministic functions of the running variable

Del Valle et al. (AEJAE 2020)

Table 2: “Impact of Fonden on Night Lights”

	Outcome	Running variable	Covariates	#Covs (not 0/1)	#Obs	Eff #Obs	Eff #Obs /#Covs	#Clusters	Covs in RD
50	difference in night lights	heavy rainfall index	characteristics of dwellings quality, health care system, education system, municipal indicators, night lights, location indicators, historic mean annual rainfall	-	2708	1563	-	1198	No

Notes: We use 24 (24 non-binary) covariates from the robustness check of Figure 4.

Table S1: Overview of the papers of the literature analysis.

	Outcome	Running variable	Covariates	#Covs (not 0/1)	#Obs	Eff #Obs	Eff #Obs /#Covs	#Clusters	Covs in RD
Londoño-Vélez et al. (AEJEP 2020)									
Table 2: "Immediate Enrollment in any Postsecondary Education, by Type of Institution"and Table A.4									
51	Immediate Enrollment in Any Postsecondary Education	SABER 11 test score	indicators for gender, age, ethic, employment status, family size, parent's education, household residential stratum, high school schedule, and private high school	21 (2)	273361	37882	1804	-	OA
52	Immediate Enrollment in Any Postsecondary Education	SISBEN wealth index		21 (2)	21071	8201	391	-	OA
Wasserman (AEJ-P&P 2021)									
Table 2: "The Effect of Losing, by Gender"									
53	Prob. Run again - Female	Margin of victory	fixed effects for state, election-year, political party, and legislative chamber (upper/lower)	92 (0)	13092	3652	121	50	Yes
54	Prob. Run again and Win - Female	Margin of victory		92 (0)	13092	3512	121	50	Yes
55	Prob. Run again - Male	Margin of victory		92 (0)	50058	12679	459	50	Yes
56	Prob. Run again and win - Male	Margin of victory		92 (0)	50058	12140	457	50	Yes

Notes: The parameter of interest is the difference of the RD estimands of female and male candidates. Here, we consider these as two separate RD regressions.

Notes: The table shows description of the respective variables ("Outcome", "Running variable", "Covariates"); the number of covariates with the number of nonbinary covariates in parentheses ("#Covs (not 0/1)"); the total sample size ("#Obs"); the number of observations within the bandwidth of the respective specification ("Eff #Obs"); the effective sample size relative to the number of covariates ("Eff #Obs / #Covs"); the number of clusters ("#Clusters"); and whether the covariates were used in the RD regression; "Yes" if they were used in the main text, "OA" if they were only used in the online appendix and "No" if they were not used ("Covs in RD").

the shortest confidence intervals among all the adjustments employing cross-fitting. The results in Table S4 are based on the robust bias correction but are otherwise analogous to the results in Table S3.

Figures S3 and S4 illustrate the asymptotic equivalence result in Theorem 1 of age as the dependent variable and bias-aware inference.[1] Specifically, they show the difference between the simulated RD estimates based on the feasible adjustments and oracle adjustments for sample sizes of 2000 and 5000. As a reference point, we also displayed the full distribution of the no covariates RD estimates. As predicted by our theory, RD estimates based on feasible and oracle adjustments are very close to each other especially compared to the distribution of no covariates RD estimates. They even become more similar when the sample size increases.

S2.3. **Additional Results for Simulation II.** Figure S5 extends the results presented in Figures 2 and 3. It presents further simulation results for bias-aware inference for age as the dependent variable and for robust bias correction for both the original outcome and age as the dependent variables. The qualitative conclusions about the bias, the standard error, and the validity of confidence intervals are very similar to the ones discussed in Section 7.3. In all cases, the simulated bias is insensitive to including many covariates, the inference based on the cross-fitted methods is valid, while the conventional linear adjustment leads to severely downward-biased standard errors and invalid confidence intervals as the number of covariates increases. The patterns in the standard deviations are different for the original outcome and age because of the different explanatory power of the covariates for these two dependent variables.

[1]The oracle and the feasible estimates are even more similar when using the original outcome as dependent variable, as the covariates do not have much explanatory power in this case. The results are also very similar when conducting inference based on robust bias correction.

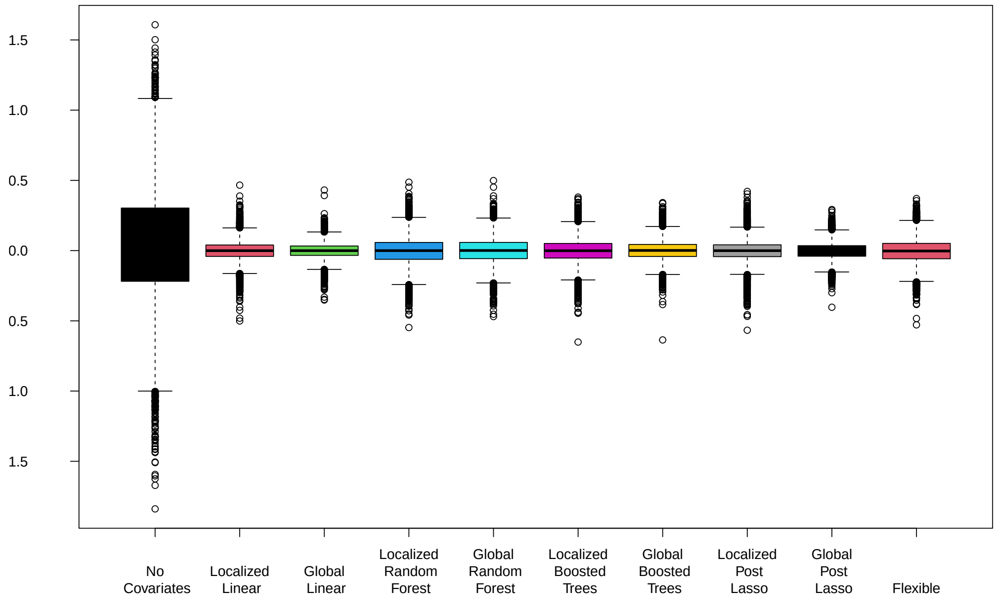

Figure S3: Difference between cross-fitted feasible and oracle estimators for $n = 2000$.

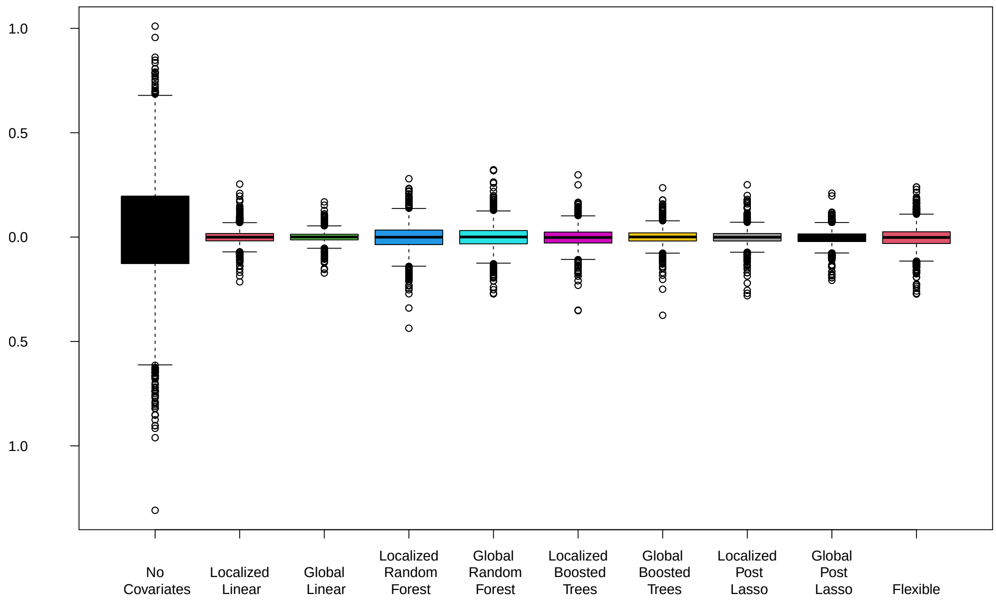

Figure S4: Difference between cross-fitted feasible and oracle estimators for $n = 5000$.

Notes: In each figure, the first box plot shows the distribution of the no covariates RD estimator and the other ones the difference of the cross-fitted feasible covariate-adjusted RD estimates and their respective oracle counterpart based on the respective adjustment methods. Simulations are based on Londoño-Vélez et al. (2020) and age is the dependent variable. See details for a description of the estimators in Section 3 for details. Results are based on 10,000 Monte Carlo draws.

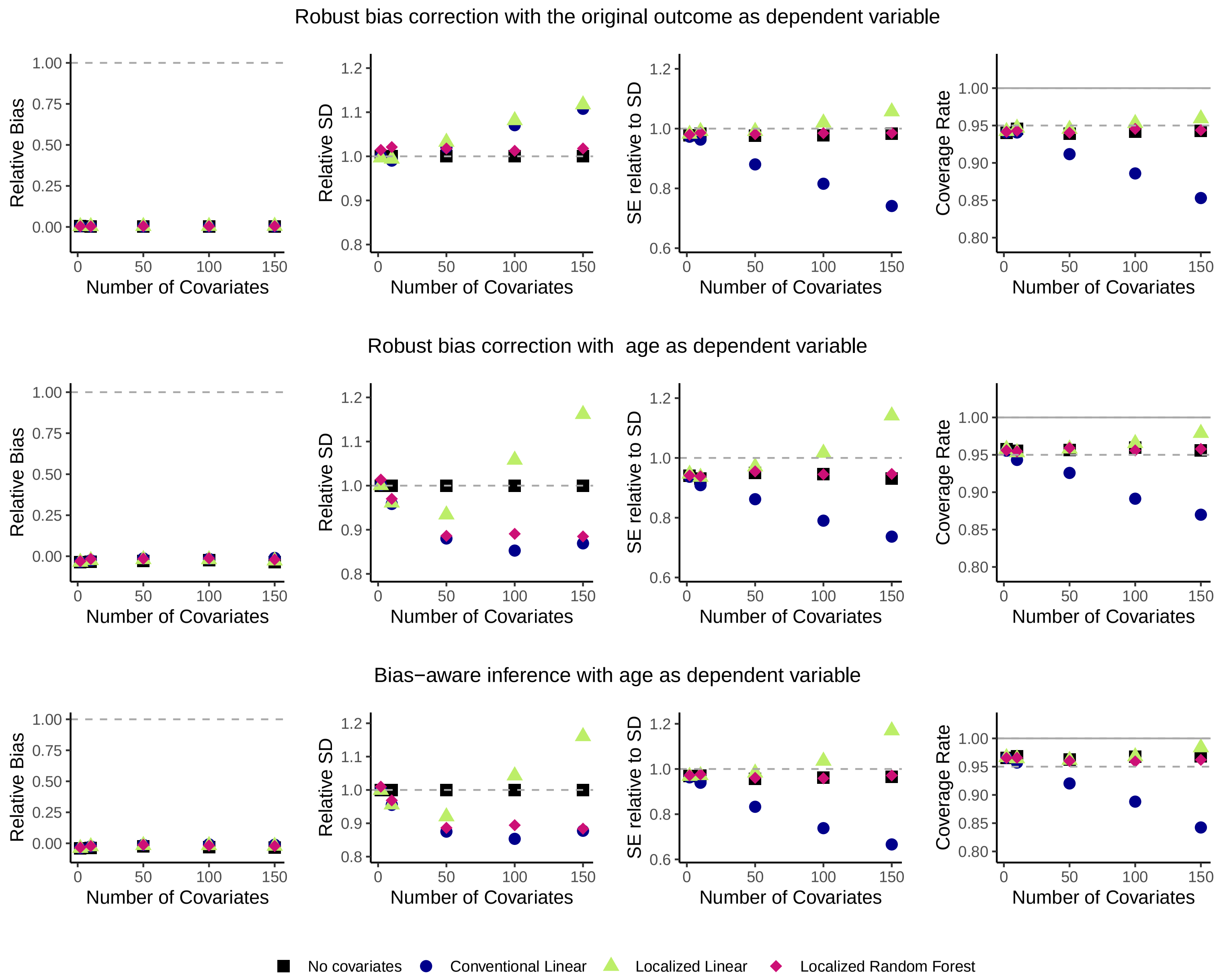

Figure S5: Additional results for Simulation II: Many covariates

Notes: Simulation design is based on Londoño-Vélez et al. (2020) and sample sizes of $n = 500$ for different numbers of covariates, see Section 7.3 for details. The three panels consider different combinations of the inference method and the dependent variable. In each row, the first and second graphs show the bias and the standard deviation of the respective estimator relative to the standard deviation of the no covariates estimator. The third graph shows the mean standard error of the respective estimator relative to its standard deviation. The last graph shows the simulated coverage of the confidence interval with 95% nominal level. See Section 3.2 and 3.3 for details of the estimators. Results are based on 10,000 Monte Carlo draws.

Table S2: Estimation results for the full restricted sample in the simulation setting of Section 7.

Adjustment Method	Bias-Aware Inference					Robust Bias Correction				
	Est x100	SE x100	Band-width	CI Length x100	CI Length % Red.	Est x100	SE x100	Band-width	CI Length x100	CI Length % Red.
Panel A - Original outcome as the dependent variable										
No Covariates	0.62	0.45	14.49	1.98	0.00	0.41	0.37	21.49	1.75	0.00
Conventional Linear	0.73	0.44	14.55	1.93	2.40	0.56	0.38	19.30	1.80	-2.58
Localized Linear	0.75	0.44	14.57	1.93	2.43	0.57	0.39	19.16	1.80	-2.92
Global Linear	0.75	0.44	14.59	1.93	2.50	0.59	0.39	19.14	1.80	-2.89
Localized Random Forest	0.68	0.44	14.43	1.93	2.17	0.50	0.38	19.31	1.79	-2.38
Global Random Forest	0.68	0.44	14.47	1.93	2.38	0.50	0.38	19.29	1.79	-2.43
Localized Boosted Trees	0.67	0.44	14.59	1.92	2.75	0.48	0.38	19.75	1.77	-1.31
Global Boosted Tree	0.67	0.44	14.50	1.93	2.38	0.49	0.38	19.83	1.77	-1.18
Localized Post-Lasso	0.71	0.44	14.72	1.91	3.09	0.55	0.38	19.56	1.78	-1.76
Global Post-Lasso	0.72	0.44	14.51	1.93	2.35	0.54	0.38	19.98	1.77	-0.87
Flexible	0.69	0.44	14.56	1.92	2.69	0.50	0.38	19.55	1.78	-1.71
Panel B - Age as the dependent variable										
No Covariates	-4.98	4.59	20.14	20.21	0.00	-7.42	5.24	15.52	23.46	0.00
Conventional Linear	-4.37	4.23	17.65	18.58	8.05	-5.77	4.49	15.69	20.23	13.77
Localized Linear	-4.39	4.24	17.65	18.60	7.95	-5.81	4.53	15.43	20.36	13.23
Global Linear	-4.05	4.24	17.62	18.64	7.77	-5.31	4.50	15.72	20.28	13.57
Localized Random Forest	-3.90	4.14	17.52	18.14	10.25	-5.10	4.29	16.14	19.25	17.96
Global Random Forest	-3.67	4.11	17.48	18.03	10.77	-4.51	4.25	16.36	19.06	18.73
Localized Boosted Trees	-4.29	4.18	17.56	18.33	9.29	-5.57	4.39	15.96	19.74	15.87
Global Boosted Tree	-4.21	4.18	17.51	18.36	9.18	-5.05	4.36	16.14	19.66	16.21
Localized Post-Lasso	-4.63	4.24	17.58	18.65	7.75	-5.91	4.52	15.49	20.32	13.38
Global Post-Lasso	-4.09	4.24	17.62	18.64	7.77	-5.36	4.50	15.71	20.28	13.54
Flexible	-3.68	4.11	17.47	18.03	10.77	-4.49	4.25	16.33	19.09	18.63

Notes: Results are based on the restricted dataset of Londoño-Vélez et al. (2020) described in Section 7. Sample size is $n = 259,419$. The columns show the estimate (Est), the standard error (SE), the bandwidth (Bandwidth), the length of confidence intervals with 95% nominal coverage (CI Length), and the percentage reduction in CI length relative to the no covariates CI length (CI Length % Red.). Estimators are described in Section 3.

Table S3: Full results for Simulation I with bias-aware inference.

Adjustment method		Original Outcome								Age							
		Mean SE x100	SD x100	Bias x100	RMSE x100	Mean Band-width	CI Cov in %	Mean CI Length x100	Mean CI Length % Red.	Mean SE x100	SD x100	Bias x100	RMSE x100	Mean Band-width	CI Cov in %	Mean CI Length x100	Mean CI Length % Red.
No Covariates		2.15	2.17	0.41	2.21	37.73	96.95	9.41	0.00	24.95	25.69	2.87	25.84	42.25	96.84	110.79	0.00
Conventional Linear		2.09	2.13	0.47	2.19	38.71	96.51	9.14	2.86	21.55	22.50	2.64	22.66	41.96	96.24	95.90	13.44
Linear Regression																	
Localized	Feasible	2.10	2.13	0.45	2.18	39.25	96.63	9.18	2.39	21.84	22.56	2.64	22.72	42.09	96.34	97.05	12.40
	Oracle	2.10	2.13	0.48	2.18	39.08	96.67	9.18	2.44	21.71	22.43	2.69	22.59	42.07	96.52	96.53	12.87
Global	Feasible	2.10	2.13	0.45	2.18	39.26	96.81	9.17	2.52	21.79	22.50	2.79	22.67	42.35	96.36	96.82	12.61
	Oracle	2.10	2.13	0.48	2.18	39.06	96.67	9.17	2.50	21.72	22.43	2.80	22.61	42.30	96.44	96.54	12.86
Random Forest																	
Localized	Feasible	2.13	2.16	0.48	2.21	40.05	96.85	9.29	1.26	21.51	22.30	2.81	22.47	42.55	96.42	95.55	13.76
	Oracle	2.10	2.13	0.43	2.17	39.26	96.75	9.18	2.45	21.10	21.97	2.96	22.17	42.61	96.18	93.80	15.33
Global	Feasible	2.13	2.16	0.48	2.21	39.50	96.72	9.28	1.31	21.42	22.21	2.94	22.40	42.93	96.36	95.16	14.11
	Oracle	2.10	2.12	0.43	2.17	39.44	96.77	9.16	2.64	20.94	21.83	2.96	22.03	42.76	96.20	93.06	16.00
Boosted Trees																	
Localized	Feasible	2.13	2.15	0.43	2.19	38.99	96.70	9.29	1.22	21.58	22.31	2.47	22.45	42.25	96.60	95.89	13.45
	Oracle	2.09	2.12	0.42	2.16	39.42	96.68	9.14	2.83	21.35	22.15	2.70	22.31	42.57	96.38	94.93	14.31
Global	Feasible	2.11	2.13	0.44	2.18	38.93	96.76	9.21	2.05	21.53	22.29	2.63	22.44	42.71	96.60	95.68	13.63
	Oracle	2.10	2.12	0.43	2.17	39.14	96.75	9.17	2.56	21.32	22.11	2.59	22.26	42.64	96.50	94.78	14.45
Post-lasso																	
Localized	Feasible	2.11	2.14	0.45	2.18	38.88	96.80	9.24	1.80	21.85	22.58	2.37	22.71	41.82	96.44	97.07	12.38
	Oracle	2.09	2.12	0.46	2.17	39.56	96.71	9.12	2.99	21.72	22.49	2.49	22.63	42.33	96.40	96.57	12.83
Global	Feasible	2.11	2.13	0.45	2.18	38.67	96.81	9.22	1.96	21.84	22.58	2.41	22.71	42.15	96.56	97.05	12.40
	Oracle	2.10	2.13	0.47	2.18	38.89	96.70	9.18	2.43	21.72	22.43	2.77	22.60	42.27	96.42	96.54	12.86
Flexible																	
	Feasile	2.11	2.13	0.44	2.18	39.03	96.69	9.19	2.28	21.35	22.10	2.65	22.25	42.25	96.48	94.87	14.37
	Oracle	2.09	2.12	0.44	2.16	39.37	96.73	9.14	2.81	20.94	21.82	2.96	22.02	42.75	96.26	93.06	16.00

Notes: Results are based on 10,000 Monte Carlo draws. The left panel shows the results for the original outcome and the right panel for age as the dependent variable based on Londoño-Vélez et al. (2020) and a sample size of $n = 5000$ (see Section 7 for details). The bandwidth is chosen and the confidence sets are constructed based on bias-aware inference. The columns show the simulated mean standard error (Mean SE), standard deviation (SD); simulated bias (Bias); root mean squared error (RMSE); average bandwidth (Mean Bandwidth), coverage of confidence intervals with 95% nominal level (CI Cov); the average confidence interval length (Mean CI Length); and the reduction in mean CI length relative to the no covariates CI length (Mean CI Length % Red.). The estimators are described in Section 3.

Table S4: Full results for Simulation I with robust bias correction.

Adjustment method		Original Outcome								Age							
		Mean SE x100	SD x100	Bias x100	RMSE x100	Mean Band-width	CI Cov in %	Mean CI Length x100	Mean CI Length % Red.	Mean SE x100	SD x100	Bias x100	RMSE x100	Mean Band-width	CI Cov in %	Mean CI Length x100	Mean CI Length % Red.
No Covariates		2.72	2.94	0.49	2.98	21.30	94.50	12.72	0.00	32.28	34.90	-5.45	35.32	21.08	94.80	150.42	0.00
Conventional Linear		2.64	2.89	0.58	2.95	21.15	93.97	12.35	2.91	27.68	29.96	-4.08	30.24	20.79	94.39	129.04	14.22
Linear Regression																	
Localized	Feasible	2.69	2.90	0.61	2.96	21.32	94.09	12.56	1.27	28.16	30.12	-4.06	30.39	21.04	94.78	131.26	12.73
	Oracle	2.67	2.87	0.60	2.94	21.30	94.21	12.45	2.13	27.87	29.80	-4.01	30.07	21.01	94.83	129.92	13.63
Global	Feasible	2.67	2.89	0.62	2.95	21.31	94.08	12.48	1.86	28.03	29.92	-3.75	30.16	21.05	94.95	130.70	13.11
	Oracle	2.67	2.88	0.61	2.94	21.30	94.22	12.45	2.11	27.93	29.81	-3.70	30.03	21.04	94.91	130.21	13.44
Random Forest																	
Localized	Feasible	2.71	2.92	0.61	2.99	21.33	94.30	12.65	0.55	27.91	29.92	-3.64	30.14	21.03	94.99	130.11	13.50
	Oracle	2.66	2.86	0.53	2.91	21.29	94.23	12.42	2.35	26.96	28.87	-3.97	29.14	21.01	94.82	125.65	
Global	Feasible	2.69	2.90	0.60	2.96	21.33	94.25	12.57	1.15	27.59	29.58	-3.48	29.78	21.05	94.80	128.62	14.50
	Oracle	2.66	2.87	0.53	2.91	21.29	94.25	12.42	2.36	26.89	28.77	-3.58	28.99	21.03	94.82	125.33	16.68
Boosted Trees																	
Localized	Feasible	2.70	2.92	0.51	2.96	21.28	94.43	12.61	0.87	27.90	29.84	-4.04	30.11	21.04	94.93	130.07	13.53
	Oracle	2.66	2.87	0.52	2.92	21.29	94.23	12.41	2.37	27.41	29.31	-4.13	29.59	21.03	94.98	127.77	15.06
Global	Feasible	2.67	2.88	0.54	2.93	21.30	94.34	12.48	1.88	27.69	29.67	-3.76	29.90	21.05	94.84	129.08	14.18
	Oracle	2.66	2.87	0.54	2.92	21.29	94.20	12.42	2.30	27.40	29.25	-3.75	29.48	21.03	94.97	127.75	15.07
Post-lasso																	
Localized	Feasible	2.69	2.91	0.52	2.96	21.31	94.57	12.58	1.08	28.12	30.16	-4.26	30.46	21.00	94.79	131.06	12.87
	Oracle	2.66	2.87	0.58	2.93	21.30	94.28	12.43	2.29	27.88	29.79	-4.21	30.09	21.01	94.82	129.95	13.61
Global	Feasible	2.68	2.89	0.54	2.93	21.30	94.41	12.49	1.76	28.06	29.99	-4.14	30.27	21.03	94.80	130.83	13.02
	Oracle	2.66	2.87	0.58	2.93	21.30	94.20	12.43	2.24	27.93	29.81	-3.75	30.04	21.04	94.94	130.22	13.43
Flexible																	
	Feasile	0.56	2.88	0.56	2.93	21.29	94.25	12.44	2.16	27.44	29.39	-3.76	29.63	21.02	94.92	127.91	14.96
	Oracle	2.66	2.87	0.54	2.92	21.29	94.24	12.41	2.42	26.89	28.77	-3.53	28.99	21.02	94.88	125.34	16.67

Notes: Results are based on 10,000 Monte Carlo draws. The left panel shows results for the original outcome and the right panel for age as the dependent variable based on Londoño-Vélez et al. (2020) and a sample size of $n = 5000$ (see Section 7 for details). The bandwidth is chosen and the confidence sets are constructed based on robust bias correction. The columns show the simulated mean standard error (Mean SE), standard deviation (SD); simulated bias (Bias); root mean squared error (RMSE); average bandwidth (Mean Bandwidth), coverage of confidence intervals with 95% nominal level (CI Cov); the average confidence interval length (Mean CI Length); and the reduction in mean CI length relative to the no covariates CI length (Mean CI Length % Red.). The estimators are described in Section 3.

REFERENCES

Akhtari, M., D. Moreira, and L. Trucco (2022): "Political turnover, bureaucratic turnover, and the quality of public services," *American Economic Review*, 112, 442–493.

Altindag, O., B. Erten, and P. Keskin (2022): "Mental health costs of lockdowns: Evidence from age-specific curfews in Turkey," *American Economic Journal: Applied Economics*, 14, 320–343.

Ambrus, A., E. Field, and R. Gonzalez (2020): "Loss in the time of cholera: Long-run impact of a disease epidemic on the urban landscape," *American Economic Review*, 110, 475–525.

Asher, S. and P. Novosad (2020): "Rural roads and local economic development," *American economic review*, 110, 797–823.

Avis, E., C. Ferraz, F. Finan, and C. Varjão (2022): "Money and politics: The effects of campaign spending limits on political entry and competition," *American Economic Journal: Applied Economics*, 14, 167–199.

Baskaran, T. and Z. Hessami (2018): "Does the election of a female leader clear the way for more women in politics?" *American Economic Journal: Economic Policy*, 10, 95–121.

Becker, S. O., I. Grosfeld, P. Grosjean, N. Voigtländer, and E. Zhuravskaya (2020): "Forced migration and human capital: Evidence from post-WWII population transfers," *American Economic Review*, 110, 1430–1463.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 82, 2295–2326.

Chin, M. (2023): "When do politicians appeal broadly? The economic consequences of electoral rules in Brazil," *American Economic Journal: Applied Economics*, 15, 183–209.

Curto-Grau, M., A. Solé-Ollé, and P. Sorribas-Navarro (2018): "Does electoral competition curb party favoritism?" *American Economic Journal: Applied Economics*, 10, 378–407.

Del Valle, A., A. de Janvry, and E. Sadoulet (2020): "Rules for recovery: Impact of indexed disaster funds on shock coping in Mexico," *American Economic Journal: Applied Economics*, 12, 164–195.

Granzier, R., V. Pons, and C. Tricaud (2023): "Coordination and bandwagon effects: How past rankings shape the behavior of voters and candidates," *American Economic Journal: Applied Economics*, 15, 177–217.

Greenstone, M., G. He, R. Jia, and T. Liu (2022): "Can technology solve the principal-agent problem? Evidence from Chinas war on air pollution," *American Economic Review: Insights*, 4, 54–70.

Johnson, M. S. (2020): "Regulation by shaming: Deterrence effects of publicizing violations of workplace safety and health laws," *American economic review*, 110, 1866–1904.

Londoño-Vélez, J., C. Rodríguez, and F. Sánchez (2020): "Upstream and downstream impacts of college merit-based financial aid for low-income students: Ser Pilo Paga in Colombia," *American Economic Journal: Economic Policy*, 12, 193–227.

Tuttle, C. (2019): "Snapping Back: Food Stamp Bans and Criminal Recidivism," *American Economic Journal: Economic Policy*, 11, 301–27.

Wasserman, M. (2021): "Up the political ladder: Gender parity in the effects of electoral defeats," *AEA Papers and Proceedings*, 111, 169–173.