Online Appendix for Socioeconomic Disparities in Privatized Pollution Remediation: Evidence from Toxic Chemical Spills

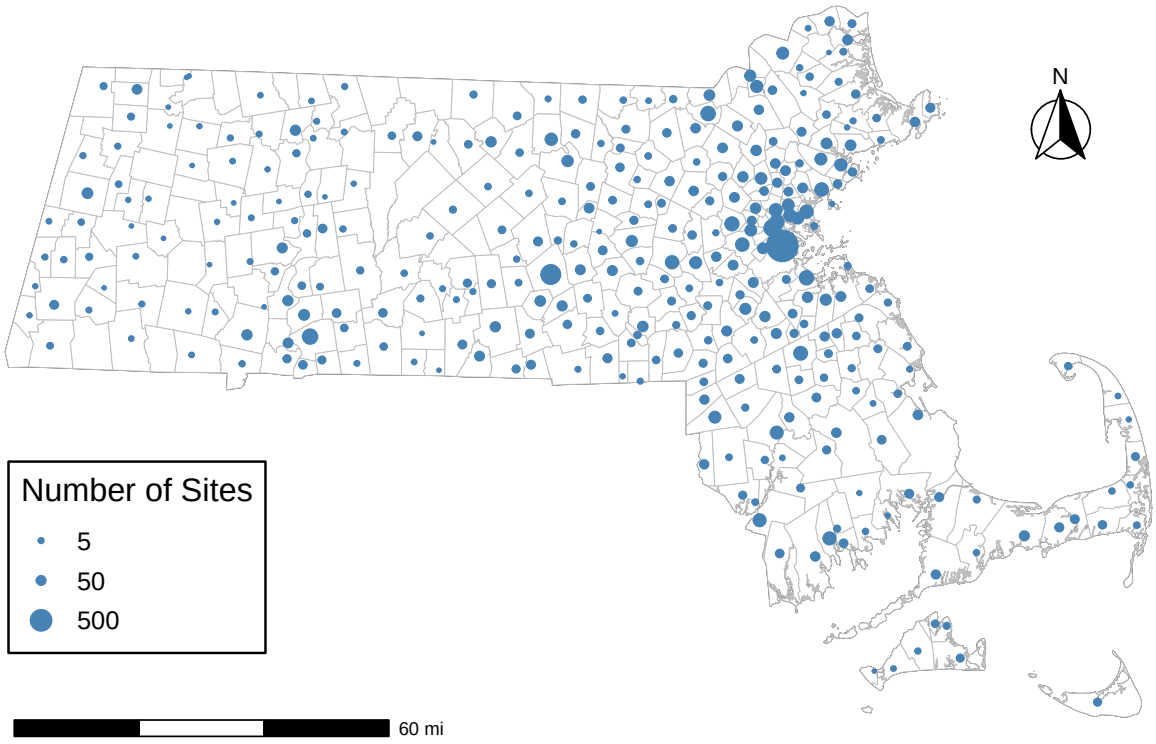
Justin Marion Jeremy West *

University of California, Santa Cruz

*Marion: marion@ucsc.edu. West: westj@ucsc.edu.

A Appendix figures and tables

Figure A1: Map of site locations across municipalities in Massachusetts



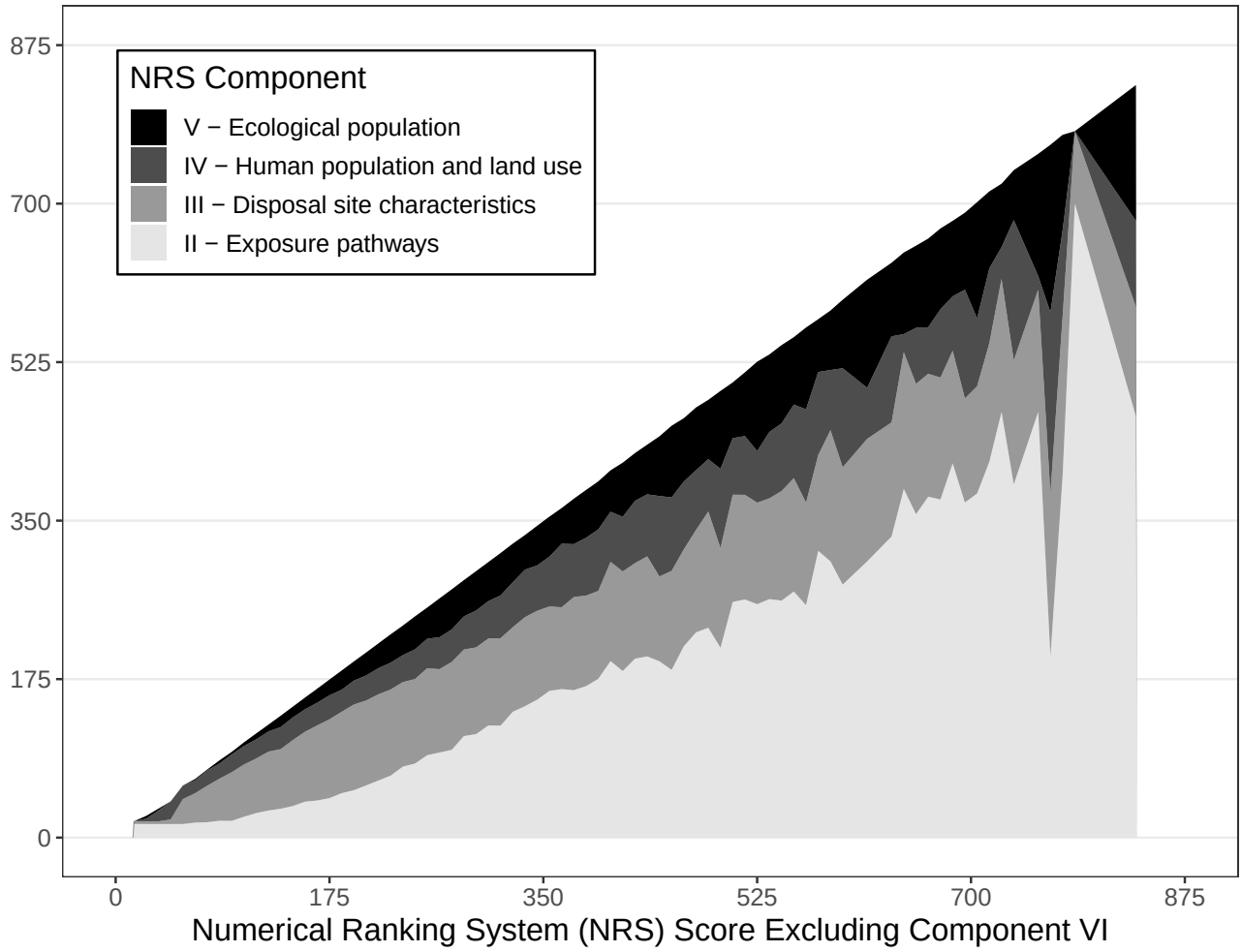
Notes: The figure shows the locations of hazardous waste sites scored using the Numerical Ranking System. The polygons show boundaries for municipalities, and the points are centered at the average coordinates of sites within each municipality. The size of the points indicates the total number of sites scored within each town.

Table A1: Numerical Ranking System components and possible score ranges

Component	Score range
I. Disposal site information	*[Not scored]*
II. Exposure pathways	*[15 – 700]*
Soil (likely presence, human exposure)	0 – 150
Groundwater (likely presence, human exposure)	0 – 150
Surface water (likely presence, human exposure)	0 – 150
Air (likely presence, affecting occupied buildings)	0 – 200
Number of sources (one, two, three or more)	0 – 50
III. Disposal site characteristics	*[3 – 180]*
Toxicity score (substance type, amount)	1 – 80
How many highly toxic substances? (none/one, more than one)	0 – 30
Substance mobility and persistence (low, medium, high)	0 – 50
Site hydrogeology (depth to groundwater, soil permeability)	2 – 20
IV. Human population and land uses	*[0 – 205]*
Population (people <0.5 mi., institutions <500ft., on-site workers)	0 – 40
Above an aquifer (no, potentially productive, or sole source)	0 – 40
Water use (proximity to public and private water supplies)	0 – 125
V. Ecological populations	*[0 – 185]*
Resource area analysis (wetlands, fish habitat, protected species)	0 – 150
Environmental toxicity analysis (substance types, concentration)	1 – 35
VI. Mitigating disposal site-specific conditions	*[± 0 – 50]*
Statutory total score range	18 – 1320
Empirical total score range	3 – 831

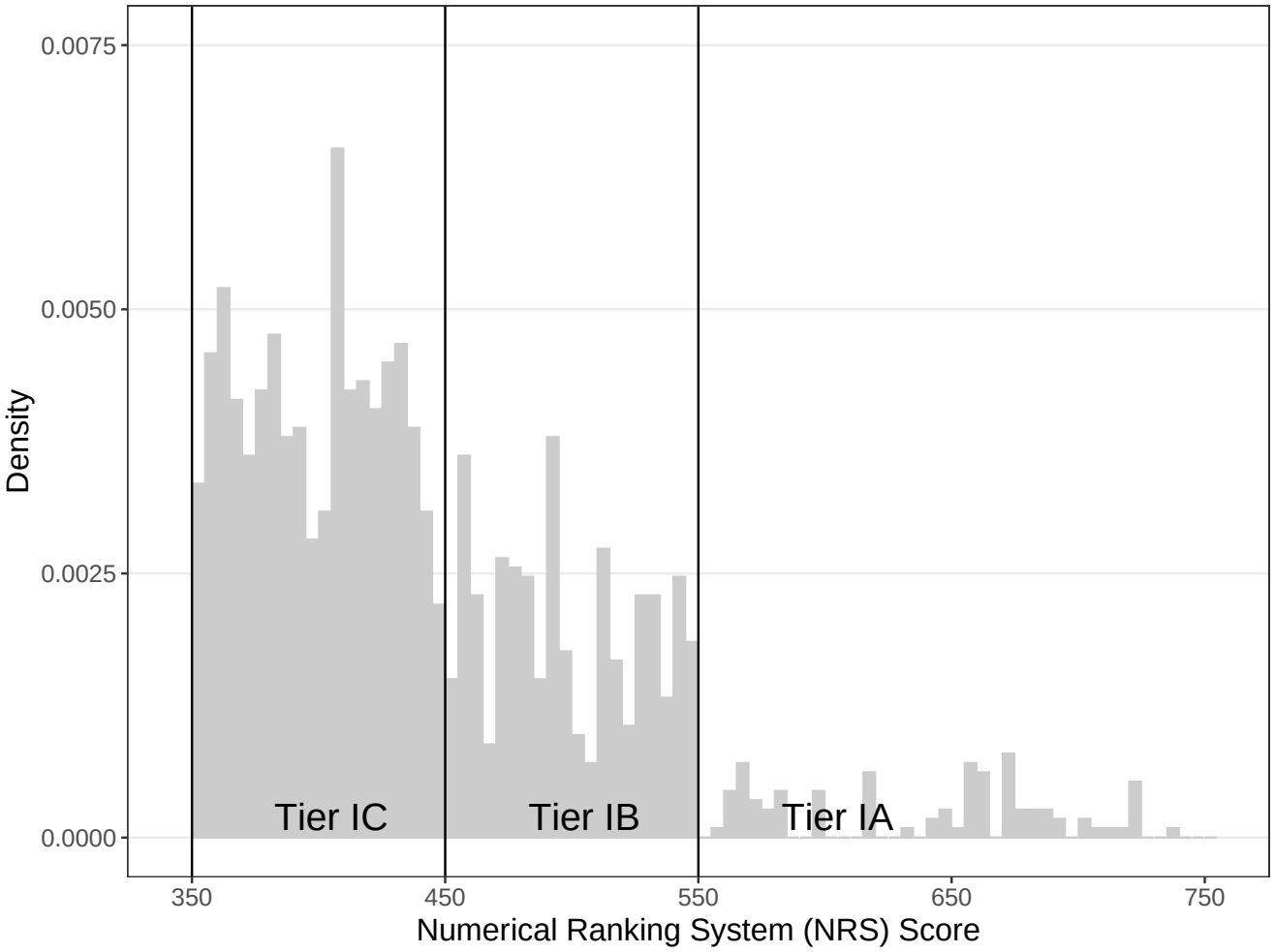
Notes: Values are sourced from the Numerical Ranking System Guidance Manual (310 CMR 40.1500). This manual of more than 80 pages is "written to assist users of the Numerical Ranking System developed by the Massachusetts Department of Environmental Protection to classify disposal sites as defined by the Massachusetts Contingency Plan and Massachusetts General Law."

Figure A2: Component contributions to total scores of sites in the Numerical Ranking System



Notes: The figure plots stacked area regions for the four component sub-scores in the Numerical Ranking System, excluding the discretionary component VI (which can take values between +/- 50 points). Note that there are very few sites with scores at the far right tail of the distribution.

Figure A3: Distribution of site scores in the NRS zoomed-in to 350-750



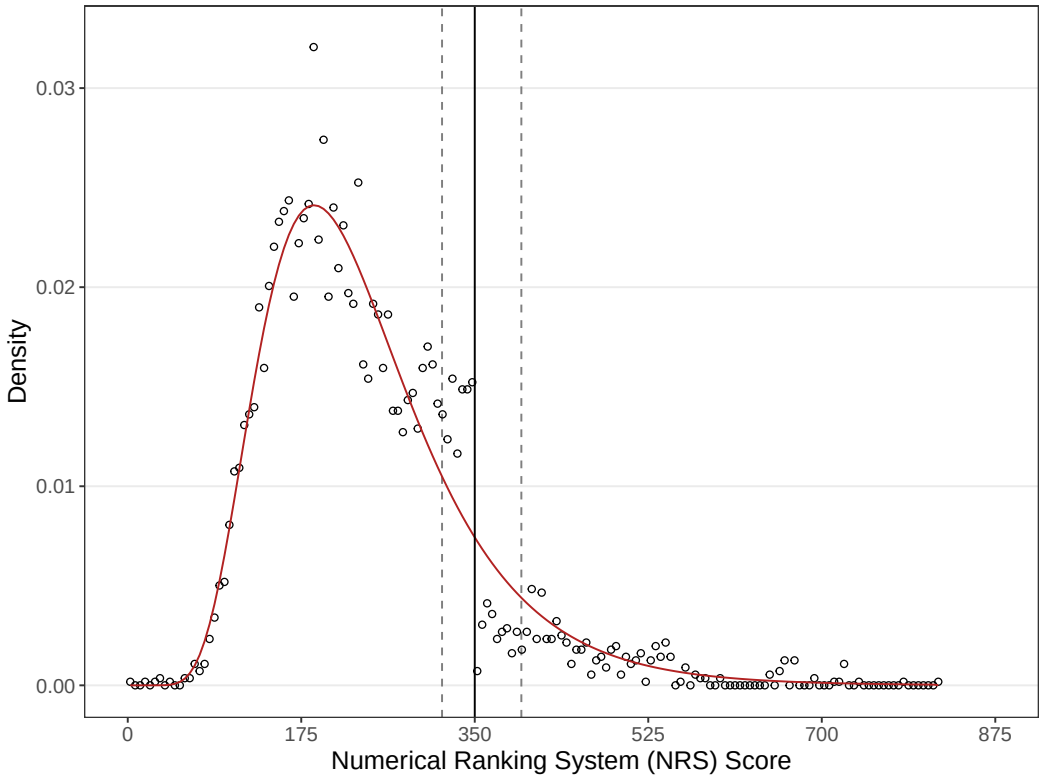
Notes: The figure plots a portion of the distribution of hazardous waste site scores in the Numerical Ranking System using a bin width of 10 points and showing the set of Tier I scores with values between 350-750. The solid vertical lines indicate the cutoffs at 350 points, 450 points, and 550 points, respectively between the Tier II/IC, Tier IC/IB, and Tier IB/IA regulatory categories.

Table A2: Predetermined neighborhood characteristics: Regression discontinuity estimates

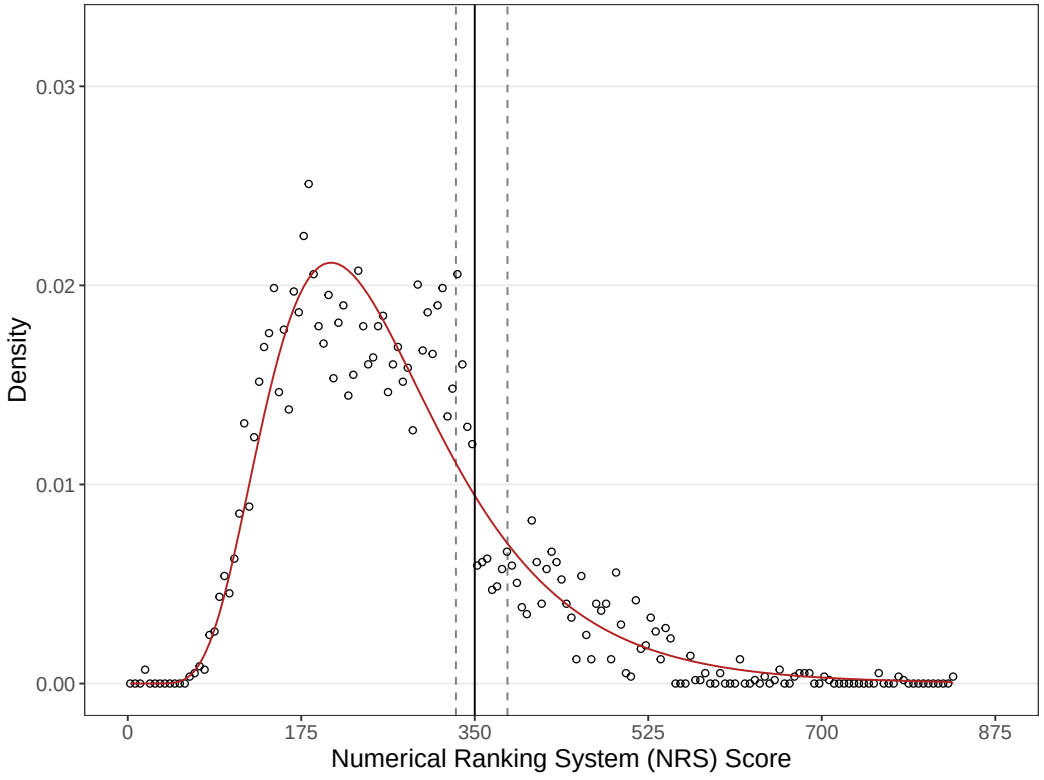
	(1)	(2)	(3)	(4)	(5)
Panel [A] Average household income ($000)					
I{Tier I}	4.841	4.58	4.656	3.048	2.754
	(1.471)	(1.361)	(1.379)	(1.170)	(1.902)
Bandwidth	66.4	65.1	58.4	69.1	50
Observations	2,918	2,898	2,605	3,101	2,184
Panel [B] Median home value ($000)					
I{Tier I}	17.77	17.22	26.78	20.89	23.91
	(7.105)	(7.520)	(6.893)	(5.408)	(11.172)
Bandwidth	73.4	65.3	58.5	72.9	50
Observations	3,184	2,847	2,561	3,129	2,153
Panel [C] White population share (%)					
I{Tier I}	5.968	5.291	2.951	0.7617	-0.3713
	(1.431)	(1.385)	(1.063)	(1.004)	(1.928)
Bandwidth	51	54.4	98.1	101.2	50
Observations	2,225	2,415	4,358	4,471	2,184
Panel [D] Adult pop. with any college (%)					
I{Tier I}	6.858	6.649	7.181	4.798	6.876
	(1.751)	(1.763)	(1.734)	(1.480)	(2.330)
Bandwidth	62.8	60.5	58.5	67.4	50
Observations	2,759	2,692	2,605	2,957	2,184
BW selection	Optimal	Optimal	Optimal	Optimal	Fixed
Year FE	No	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes
County FE	No	No	No	Yes	Yes

Notes: Each column presents results from a separate regression discontinuity estimation for how the outcome in each panel varies where crossing the Tier II to Tier I threshold at 350 total points in the Numerical Ranking System. All regressions use the "rdrobust" software package developed and provided by Calonico et al. (2014). Heteroskedasticity-robust bias-corrected standard errors are selected using the same package, as are optimal bandwidths using a triangular kernel. Where included, tier-assignment year FE are fixed effects for each year (1994-2013) of NRS site scoring, and region FE are fixed effects for each of the four MassDEP office regions.

Figure A4: Estimated manipulation regions and counterfactual densities for subsamples of sites in below and above median Census Tracts based on average household earned income



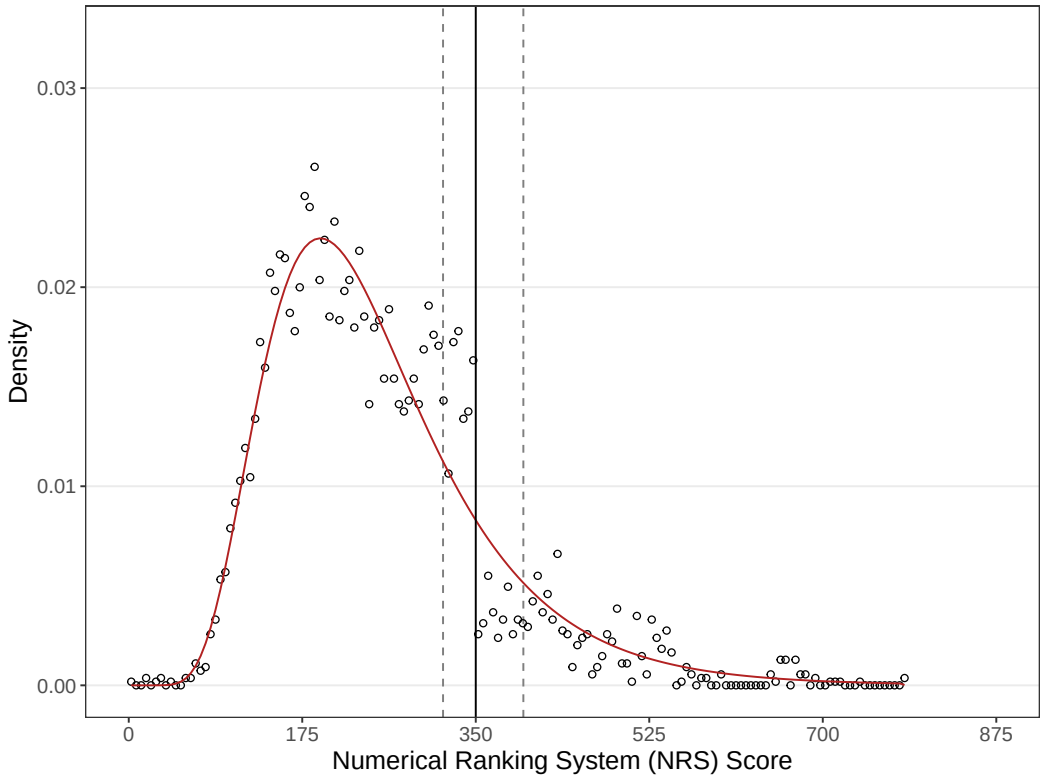
(a) Below median Census Tracts



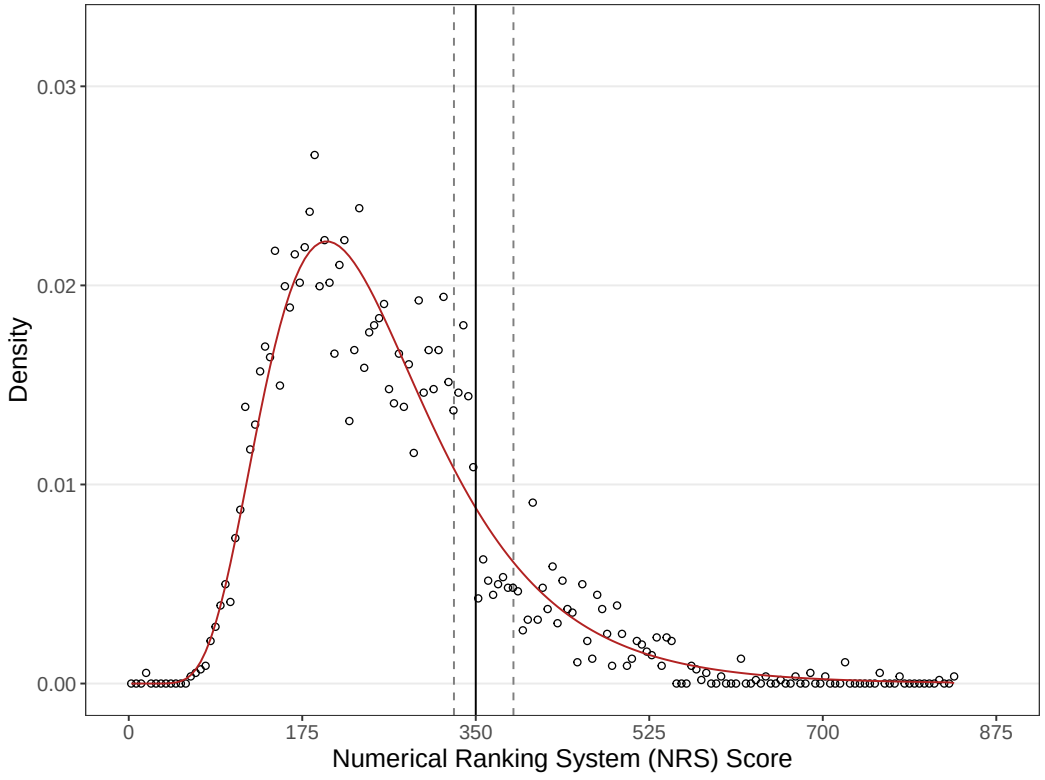
(b) Above median Census Tracts

Notes: The figures plot the score distributions for the indicated subsamples, the estimated regions over which score manipulation is present, and the counterfactual density functions.

Figure A5: Estimated manipulation regions and counterfactual densities for subsamples of sites in below and above median Census Tracts based on median home value



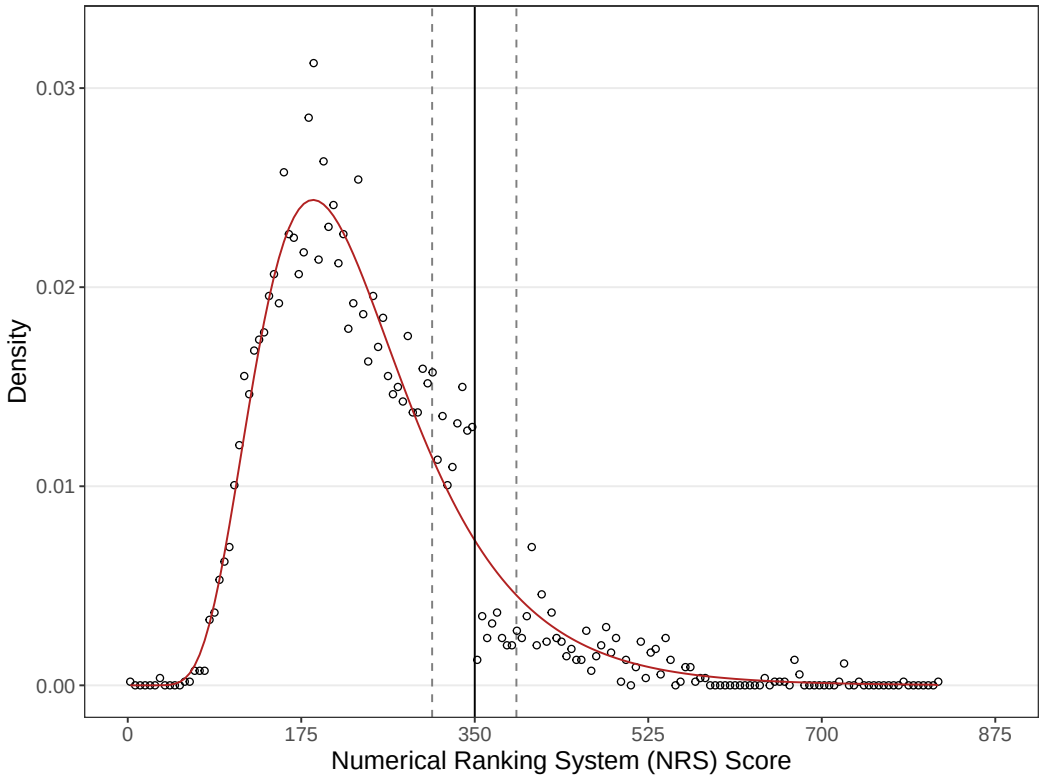
(a) Below median Census Tracts



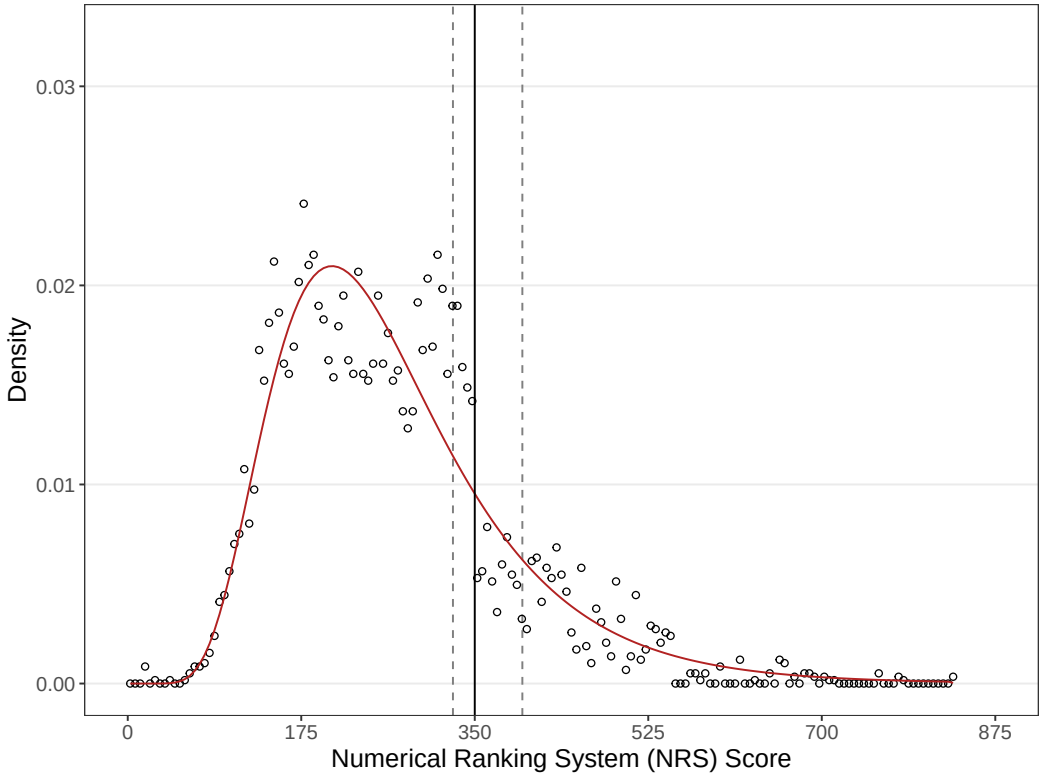
(b) Above median Census Tracts

Notes: The figures plot the score distributions for the indicated subsamples, the estimated regions over which score manipulation is present, and the counterfactual density functions.

Figure A6: Estimated manipulation regions and counterfactual densities for subsamples of sites in below and above median Census Tracts based on white population share



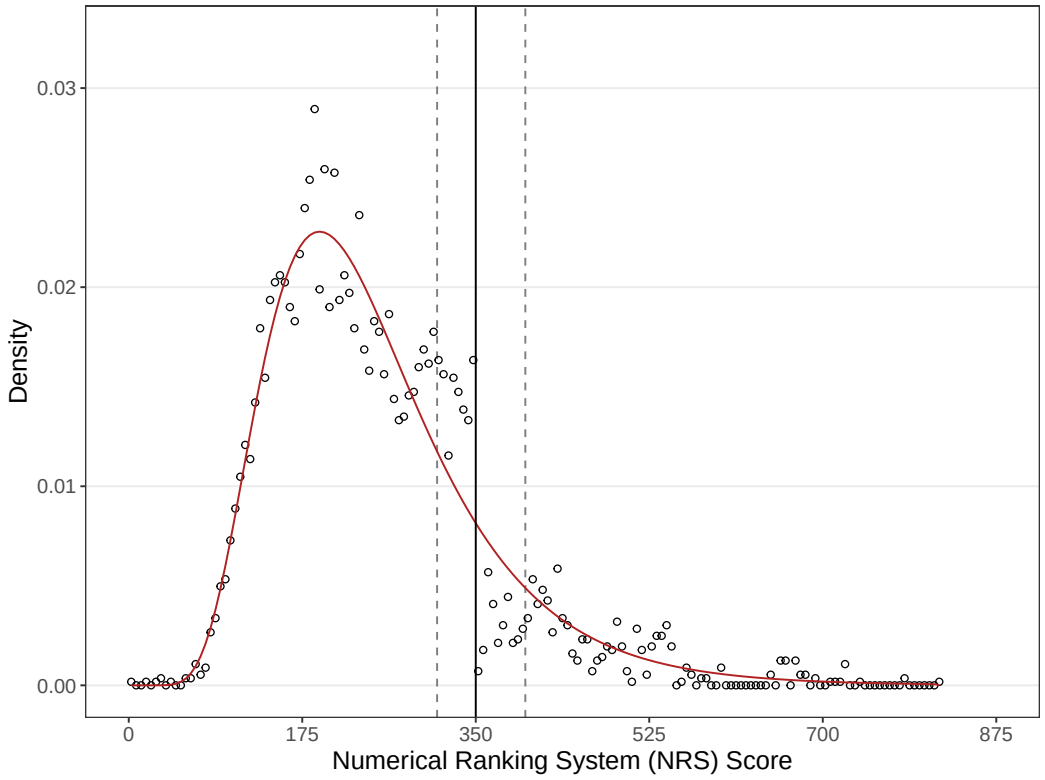
(a) Below median Census Tracts



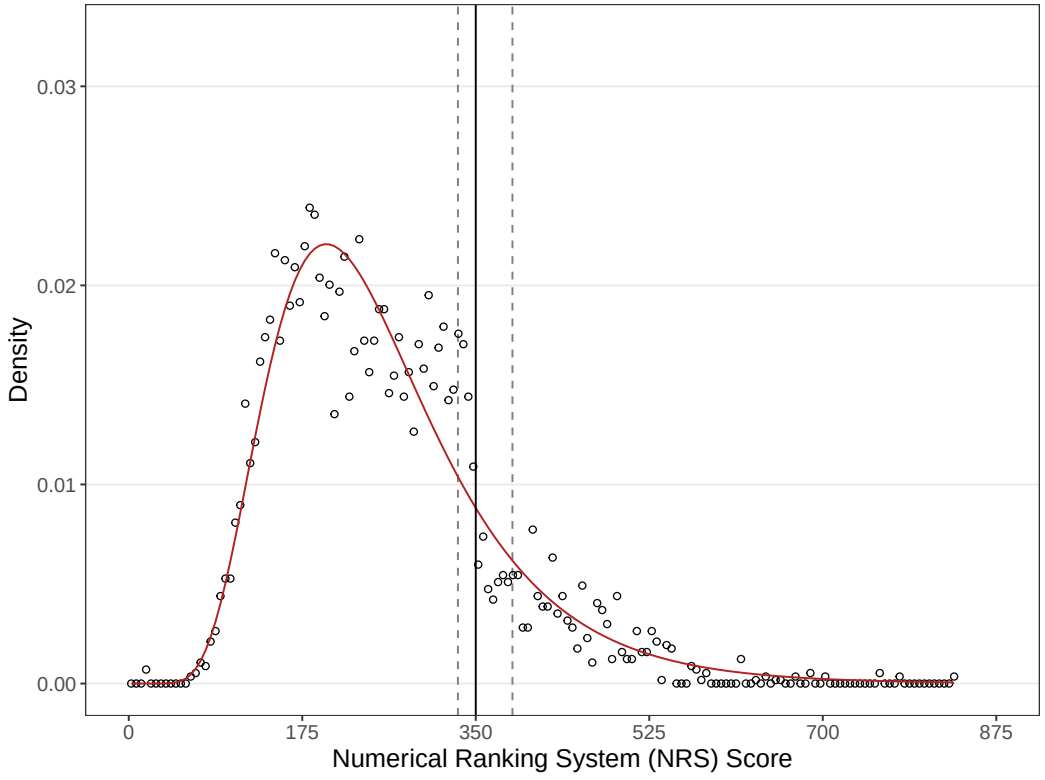
(b) Above median Census Tracts

Notes: The figures plot the score distributions for the indicated subsamples, the estimated regions over which score manipulation is present, and the counterfactual density functions.

Figure A7: Estimated manipulation regions and counterfactual densities for subsamples of sites in below and above median Census Tracts based on adult population with any college



(a) Below median Census Tracts



(b) Above median Census Tracts

Notes: The figures plot the score distributions for the indicated subsamples, the estimated regions over which score manipulation is present, and the counterfactual density functions.

Table A3: Relationship between the estimated excess density at each site's NRS score and neighborhood characteristics: Linear regression estimates

	Dep. variable: Excess density X 1000							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Household earned income	−0.408				−0.074	−0.076	−0.108	−0.433
	(0.080)				(0.134)	(0.135)	(0.138)	(0.167)
Median home value		−0.233			0.038	0.032	−0.655	−0.547
		(0.081)			(0.133)	(0.133)	(0.160)	(0.169)
White population share			−0.673		−0.611	−0.605	−0.324	−0.186
			(0.079)		(0.088)	(0.089)	(0.094)	(0.104)
Adult pop. with any college				−0.333	−0.136	−0.128	0.211	0.409
				(0.082)	(0.144)	(0.144)	(0.152)	(0.165)
Dep. variable mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year fixed effects	No	No	No	No	No	Yes	Yes	Yes
Region fixed effects	No	No	No	No	No	No	Yes	Yes
County fixed effects	No	No	No	No	No	No	No	Yes
Observations	11,064	11,064	11,064	11,064	11,064	11,064	11,064	11,064

Notes: Each column presents results from a linear regression of the estimated excess density at a site's NRS score on the four 1990 Census Tract covariates, expressed as percentiles within the state. Heteroskedasticity-robust standard errors are in parentheses. Where included, the year fixed effects are for the year of tier assignment and region fixed effects are for each of the four MassDEP office regions.

B Methodological appendix

This appendix section provides additional details and discussion about the bunching estimator that we use to estimate the width of the manipulation region and the counterfactual density function. As described in Section III (Results) of the paper, our estimator is adapted from Diamond and Persson (2016), Kleven (2016), and Chen et al. (2021). The methodology uses k-fold cross-validation with a grid search over possible widths of the manipulation region. After the data-driven approach selects a manipulation region, we then use the full sample of data outside of the chosen manipulation region to estimate the counterfactual log-normal distribution. We compare the observed data to this estimated counterfactual to quantify manipulation, using a bootstrap procedure for inference.

To recover the unmanipulated distribution and the width of the manipulation region, we first collapse the data of Numerical Ranking System (NRS) scores into the density at each score value. The domain of empirical score values spans from three to 831 points. Let the density at score value s be defined as d_s, which we model as:

$$\underbrace{d_s}_{\substack{\text{Observed}\\ \text{density}}} = \underbrace{\Phi(\theta, s)}_{\substack{\text{Unmanipulated}\\ \text{distribution}}} + \underbrace{\sum_{j=\underline{s}}^{349} \gamma_j \cdot 1[s=j]}_{\substack{\text{Excess}\\ \text{density}}} - \underbrace{\sum_{j=350}^{\overline{s}} \gamma_j \cdot 1[s=j]}_{\substack{\text{Missing}\\ \text{density}}} + \underbrace{\epsilon_s}_{\substack{\text{Sampling}\\ \text{error}}} \quad (1)$$

The counterfactual density at each score value is obtained as the predicted value of Equation (1) omitting the contribution of the dummies in the excluded region $[\underline{s}, \overline{s}]$, i.e. $\hat{d}_s = \Phi(\hat{\theta}, s)$. Excess (missing) mass is then estimated as the difference between the observed and counterfactual density for each score value in $[\underline{s}, \overline{s}]$.[1] We specify that the unmanipulated distribution Φ is log-normal, parameterized by mean μ and standard deviation σ.[2]

We determine the manipulation region using k-fold (k=5) cross-validation with a grid search over all possible combinations of $\underline{s} \in [300, 345]$ and $\overline{s} \in [355, 400]$.[3] For each guess of the manipulation region, $[\underline{s}, \overline{s}]$, we use a constrained optimization by linear approximations

[1] As discussed in the main text, the NRS also has tier thresholds at 450 and 550 points, however, there is very little density in the right tail of the score distribution. Less than five percent of sites are scored above 450 points. Our estimator focuses only on the manipulation region around the 350 point Tier I/II threshold.

[2] Bunching estimators in the literature often specify the unmanipulated density function as a linear combination of polynomial basis functions. In our setting, this leads to much worse quality-of-fit and implausible estimated manipulation regions, compared to imposing log-normal structure for the density function.

[3] The grid search is computationally intensive and run-time scales exponentially with the size of the grid, so our primary procedure constrains the possible manipulation region to be within 50 points of the threshold. This also reduces the bias from any manipulated mass around the Tier IC/IB 450-point threshold. We confirmed that the estimated manipulated density is unchanged if we further expand the grid size.

(COBYLA) direct search algorithm to estimate the counterfactual density. Using only score values outside of $[\underline{s}, \overline{s}]$, the nonlinear optimization solves for the log-normal parameters μ and σ that minimize the sum of squared errors between the counterfactual density and the observed density:

$$\underset{\{\mu,\sigma\}}{\text{Min}} \sum_{s \notin [\underline{s},\overline{s}]} (d_s - \Phi(\mu, \sigma, s))^2 \tag{2}$$

This estimation is done using the 80 percent training sample.[4] We then calculate the out-of-sample mean squared error (MSE) for the 20 percent hold-out sample using the estimated $\hat{\mu}$ and $\hat{\sigma}$. This is done separately for each of the five folds, and then we average the MSE over the five folds for each combination of $[\underline{s}, \overline{s}]$. We then select the manipulation region, $[\underline{s}, \overline{s}]$, that yields the smallest out-of-sample MSE, conditional on passing a statistical test that total excess mass equals total missing mass. To operationalize this statistical test, we pool the five folds and calculate the residuals, $d_s - \Phi(\hat{\mu}, \hat{\sigma}, s)$, using the estimated $\hat{\mu}$ and $\hat{\sigma}$. Our test criteria is that the absolute value of the total prediction error in the manipulation region, $\sum_{\underline{s}}^{\overline{s}}(d_s - \Phi(\hat{\mu}, \hat{\sigma}, s))$, is smaller than the 10th percentile of the absolute value of these residuals. In practice, the selected manipulation region has a difference in total excess mass and total missing mass that is very close to zero (≈ 0.001).

Once we have determined the out-of-sample MSE-minimizing manipulation region, $[\underline{s}, \overline{s}]$, and counterfactual distribution, $\Phi(\hat{\mu}, \hat{\sigma})$, we compute the total missing density in $[350, \overline{s}]$ using Equation (1). Likewise, we compute the share of scores in $[350, \overline{s}]$ that are manipulated by dividing the total estimated missing density by the total counterfactual density in $[350, \overline{s}]$.

We bootstrap standard errors for the estimated missing density and share of scores manipulated by drawing score values with replacement from the original score distribution and collapsing to score density, and then re-estimating Equation (2) and calculating missing density using Equation (1) for each bootstrap sample, with 1000 repetitions.

References

Z. Chen, Z. Liu, J. C. S. Serrato, and D. Y. Xu. Notching R&D investment with corporate income tax cuts in China. *American Economic Review*, 111(7):2065–2100, 2021.

R. Diamond and P. Persson. The long-term consequences of teacher discretion in grading of high-stakes tests. NBER Working Paper w22207, 2016.

[4]Following the literature, when randomly binning the data into five groups for cross-validation, we sample from the score density values, d_s, instead of sampling from the uncollapsed data on waste sites.

H. J. Kleven. Bunching. *Annual Review of Economics*, 8(1):435–464, 2016.