Online Appendix

Environmental Disadvantage and the Distribution of Regulatory Inspections in the United States

Contents

A Data Construction and Measurement **3**

A.1 Sample Construction 3

A.2 Environmental Inspection Data 4

A.3 Pollution Burden 5

A.4 Disadvantaged County Classification 6

A.5 Facility Counts 6

A.6 Data Merging and Geographic Harmonization 7

B Descriptive Statistics and Geographic Patterns **7**

B.1 Summary Statistics 7

B.2 Distribution of Inspections 8

B.3 Geographic Distribution 9

B.4 County-Year Panel Coverage 10

B.5 Distribution of Pollution Burden 12

B.6 Pollution Burden and Inspection Activity 13

C Additional Model Results **14**

C.1 Full Negative Binomial Results 14

C.2 Air Inspection Model 15

C.3 Water Inspection Model 16

D Panel and Sample Robustness **17**

D.1 Balanced-Panel Results 17

D.2 Full-Sample and Geographic-Sample Comparisons 19

D.3 Extreme-Value Sensitivity 20

D.4 Continuous Disadvantage Measure 21

D.5 Subperiod Stability of the Disadvantage Estimate 23

D.6 States with Within-State Variation 25

D.7 Robustness Controlling for County Poverty Rate 27

E Alternative Facility Specification **28**

E.1 Facility-Offset Model 28

F Count-Model Diagnostics **30**

F.1 Overdispersion and Model Fit 30

G Program-Specific Predicted Inspections **31**

G.1 Predicted Air Inspections 31

G.2 Predicted Water Inspections 32

Appendix A. Data Construction and Measurement

A.1 Sample Construction

The unit of analysis is the county-year, with annual observations covering 2013 through 2022. The initial merged dataset contained 23,638 county-year observations. During data cleaning, seven duplicate county-year records were identified and resolved, resulting in 23,631 unique county-year observations.

The primary analysis is restricted to counties and county-equivalent units in the 50 states and the District of Columbia. Observations from American Samoa, Guam, the Northern Mariana Islands, Puerto Rico, and the U.S. Virgin Islands are excluded. The resulting primary analytic sample contains 23,155 county-year observations representing 2,500 counties and county-equivalent units.

The panel is unbalanced because not every county is observed in every year from 2013 through 2022. The primary analysis retains all county-year observations containing the information required for estimation. This approach preserves geographic coverage and avoids discarding usable observations from counties with incomplete annual coverage.

As a robustness check, the primary model is re-estimated using a balanced panel containing only counties observed in all ten years. The balanced-panel sample contains 21,500 county-year observations from 2,150 counties and is reported in Appendix D.1.

County observations are linked across datasets using Federal Information Processing Standards county codes. The final analytic file combines annual measures of environmental inspection activity, modeled pollution burden, regulated facility concentration, and county disadvantage status. The construction and measurement of each component are described below.

A.2 Environmental Inspection Data

Environmental inspection data come from the U.S. Environmental Protection Agency's Enforcement and Compliance History Online (ECHO) database. ECHO integrates compliance and enforcement information reported to EPA national data systems by federal, state, local, and tribal environmental agencies. The database includes information on regulated facilities, permits, compliance evaluations and inspections, violations, enforcement actions, and penalties (U.S. Environmental Protection Agency, 2025a). ECHO therefore provides a national source for examining how environmental regulatory activity is distributed across facilities and geographic areas.

This study uses annual inspection records associated with facilities regulated under major federal environmental programs. ECHO maintains program-specific compliance and enforcement information for facilities subject to statutes including the Clean Air Act (CAA) and Clean Water Act (CWA), while preserving information on the location of regulated facilities and the dates of compliance-monitoring activities (U.S. Environmental Protection Agency, 2025a). Inspection records are aggregated from the facility level to the county-year level using county FIPS codes, producing annual counts of regulatory monitoring activity within each county.

Three dependent variables are constructed from the ECHO inspection records and aggregated to the county-year level. Table A1 summarizes their definitions.

Table A1: Environmental Inspection Outcomes

Outcome	**Definition**	**Regulatory Program**
Total inspections	Annual number of environmental inspections recorded for facilities within a county across the air and water programs included in the analysis.	Clean Air Act and Clean Water Act
Air inspections	Annual number of inspections recorded for facilities within a county under the air regulatory program.	Clean Air Act
Water inspections	Annual number of inspections recorded for facilities within a county under the water regulatory program.	Clean Water Act

Note: All outcomes are annual county-level counts constructed from EPA Enforcement and Compliance History Online (ECHO) inspection records.

A.3 Pollution Burden

Pollution burden is measured using county-level scores from the U.S. Environmental Protection Agency's Risk-Screening Environmental Indicators (RSEI) model. RSEI combines information on chemical toxicity, modeled exposure, and the potentially exposed population to provide a screening-level measure of relative pollution-related risk (U.S. Environmental Protection Agency, 2025b). Higher scores indicate greater potential risk-related impacts, but the measure should not be interpreted as a direct estimate of individual health risk or realized health outcomes.

The RSEI score is highly right-skewed, so the analysis uses the transformation

$$\text{Logged Pollution Burden}_{it} = \log(1 + \text{RSEI Score}_{it}), \tag{A.1}$$

where i indexes counties and t indexes years. Adding one retains observations with zero RSEI scores, while the logarithmic transformation reduces the influence of extremely large values. The transformed measure is mean-centered in the interaction models, so the coefficient on disadvantaged status represents the difference in expected inspection activity at the average level of logged pollution burden.

A.4 Disadvantaged County Classification

Community disadvantage is measured using the Climate and Economic Justice Screening Tool (CEJST), which identifies disadvantaged census tracts using indicators of socioeconomic and environmental burden (White House Council on Environmental Quality, 2024). Because the analysis is conducted at the county-year level, tract-level classifications are aggregated to counties.

A county is coded as disadvantaged if it contains at least one census tract designated as disadvantaged by CEJST; all other counties are coded as non-disadvantaged. This binary measure captures the presence of disadvantaged communities within a county but does not distinguish between counties containing one disadvantaged tract and those containing several disadvantaged tracts.

A.5 Facility Counts

Facility concentration is measured as the annual number of regulated facilities observed within each county and is included to account for differences in the number of regulated sites that may generate inspection activity. Counties with more regulated facilities generally provide more opportunities for inspections, making facility concentration an important control in models of county-level inspection counts.

The distribution of facility counts is right-skewed, so the primary specification uses the natural logarithm of the number of facilities to reduce the influence of extremely large values and better represent proportional differences across counties. A separate offset specification,

reported in Appendix E.1, imposes a one-to-one proportional relationship between facilities and expected inspections as an alternative benchmark.

A.6 Data Merging and Geographic Harmonization

The final analytic dataset is constructed by linking inspection activity, pollution burden, facility counts, and disadvantage classifications using county Federal Information Processing Standards (FIPS) codes as the common geographic identifier. Annual ECHO inspection and facility records and annual RSEI burden measures are aligned at the county-year level, while tract-level CEJST classifications are first aggregated to the county level using the procedure described in Appendix A.4 and then merged with the annual data.

Geographic harmonization ensures that county and county-equivalent units are consistently identified across the underlying data sources before estimation. The resulting merged panel retains the county-year structure used throughout the analysis, with one observation representing a county in a given year.

Appendix B. Descriptive Statistics and Geographic Patterns

B.1 Summary Statistics

Table B1 reports descriptive statistics for the primary analytic sample, including annual total, air, and water inspections, pollution burden, facility concentration, chemical coverage, and disadvantaged-county status.

Table B1: Descriptive Statistics for the Main County-Year Sample

Variable	N	Mean	SD	Median	Min	Max
Total inspections	23155	47.005	83.781	23.000	0	1571.00
Air inspections	23155	19.541	46.116	6.000	0	1571.00
Water inspections	23155	27.464	54.053	13.000	0	1178.00
Pollution burden	23155	161099.684	1493226.423	1155.870	0	83594637.66
Logged pollution burden	23155	6.663	4.132	7.053	0	18.24
Regulated facilities	23155	8.181	16.672	4.000	1	384.00
Chemicals	23155	16.241	18.518	11.000	1	368.00
Disadvantaged county	23155	0.886	0.318	1.000	0	1.00

Notes: The unit of analysis is the county-year. The primary sample contains 23,155 observations from 2,500 counties and county-equivalent units in the 50 states and the District of Columbia from 2013 through 2022. Total inspections equal the sum of annual air and water inspections. Logged pollution burden equals $\log(1 + \text{RSEI score})$.

B.2 Distribution of Inspections

Figure B1 shows the distribution of annual total inspections for disadvantaged and non-disadvantaged county-year observations. Inspection counts are transformed using $\log(1 + \text{inspections})$ to reduce the influence of the long right tail. The distribution for disadvantaged counties is shifted toward higher inspection levels, although substantial overlap remains between the two groups. The figure is descriptive and does not adjust for pollution burden, facility concentration, state differences, or year effects.

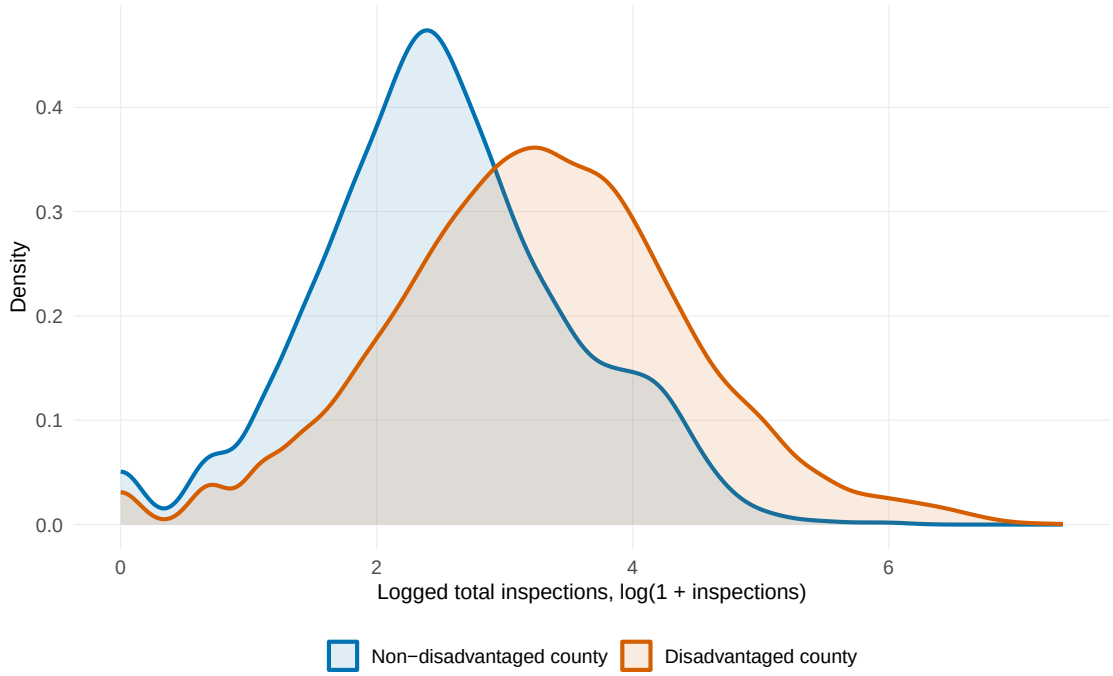


Figure B1: **Distribution of Annual Total Inspections.** The density curves compare disadvantaged and non-disadvantaged county-year observations in the primary sample.

B.3 Geographic Distribution

Figure B2 shows the geographic distribution of disadvantaged counties across the primary analytic sample. The map illustrates the spatial distribution of the county-level disadvantage classification used throughout the analysis.

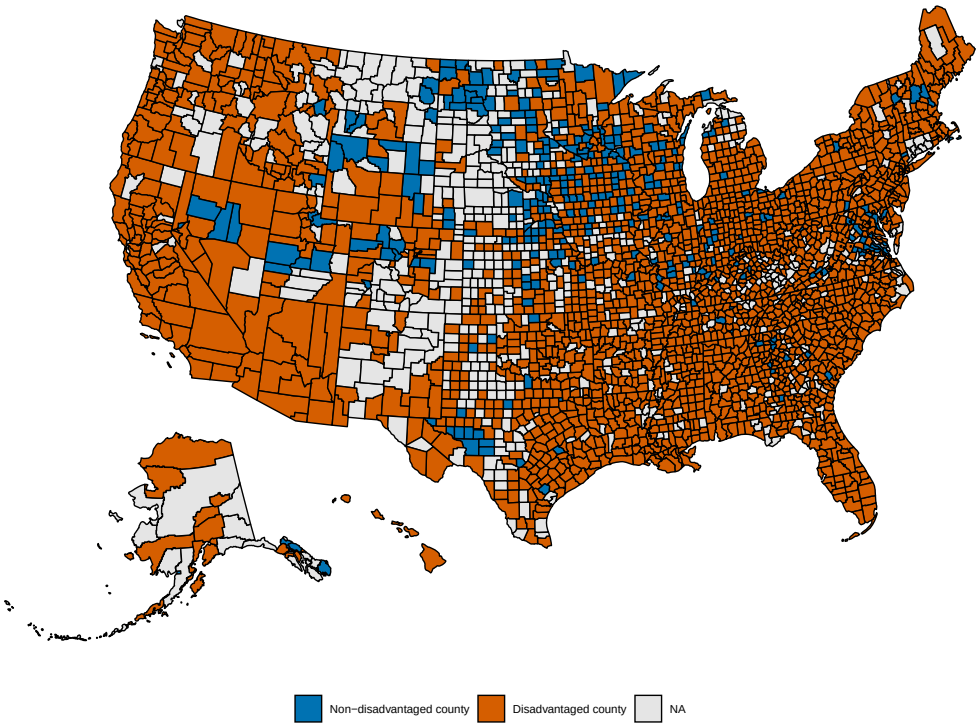


Figure B2: **Geographic Distribution of Disadvantaged Counties.** A county is classified as disadvantaged when it contains at least one census tract designated as disadvantaged by the Climate and Economic Justice Screening Tool.

B.4 County-Year Panel Coverage

The primary dataset forms an unbalanced county-year panel, with some counties observed for fewer than the full ten years between 2013 and 2022. Table B2 reports the number of counties according to the number of years observed, while Table B3 reports annual county coverage across the study period. Counties observed in all ten years form the balanced-panel sample used in the robustness analysis reported in Appendix D.1.

Table B2: County Coverage Across the Ten-Year Panel

Years observed	Counties
1	70
2	38
3	24
4	35
5	39
6	34
7	35
8	22
9	53
10	2150

Notes: The table groups counties and county-equivalent units by the number of years observed between 2013 and 2022. Counties observed in all ten years form the balanced-panel robustness sample.

Table B3: Number of Counties Observed by Year

Year	Counties observed
2013	2312
2014	2318
2015	2309
2016	2314
2017	2322
2018	2314
2019	2317
2020	2303
2021	2306
2022	2340

Notes: The primary sample includes counties and county-equivalent units in the 50 states and the District of Columbia. Annual coverage remains broadly stable across the 2013–2022 period.

B.5 Distribution of Pollution Burden

Figure B3 shows the distribution of logged pollution burden for disadvantaged and non-disadvantaged county-year observations. The distribution for disadvantaged counties is shifted toward higher pollution-burden values, although substantial overlap remains between the two groups. The figure provides a descriptive comparison of environmental burden and does not account for facility concentration, state differences, year effects, or other covariates used in the regression models.

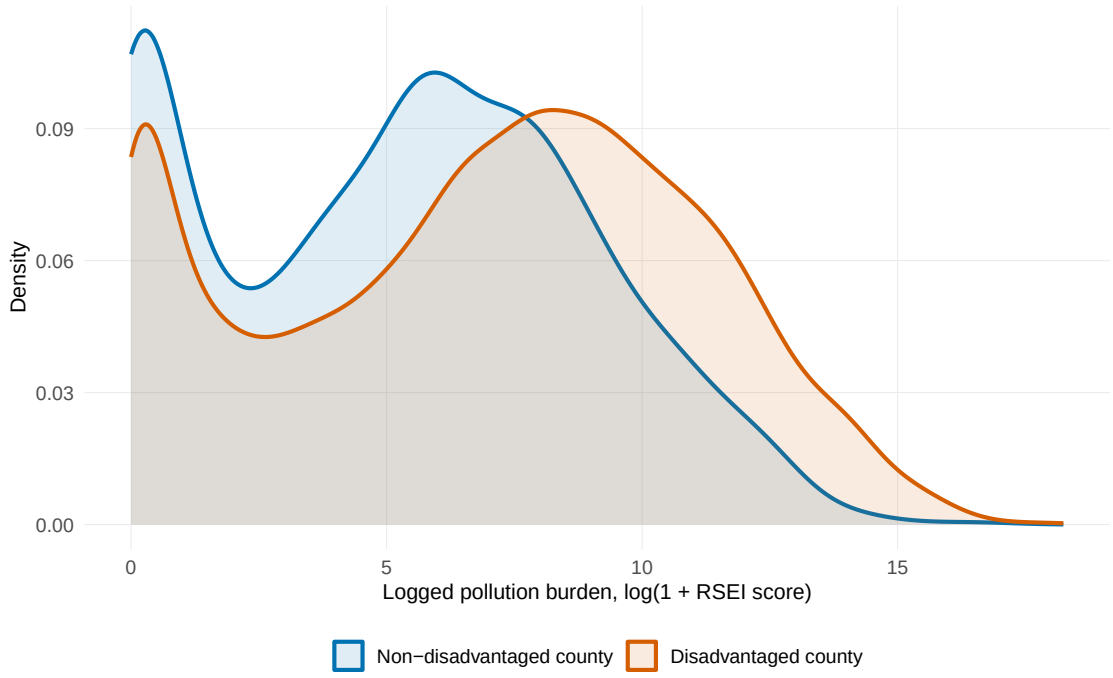


Figure B3: **Distribution of RSEI Pollution Burden.** The density curves compare disadvantaged and non-disadvantaged county-year observations.

B.6 Pollution Burden and Inspection Activity

Figure B4 presents the descriptive relationship between logged pollution burden and annual total inspections separately for disadvantaged and non-disadvantaged county-year observations. Logged pollution burden is divided into quantile-based bins, and each point reports the mean pollution burden and mean number of inspections within a bin. Inspection activity generally increases with pollution burden, while disadvantaged counties exhibit higher mean inspection counts across much of the distribution. The figure is descriptive and does not adjust for facility concentration, state differences, or year effects.

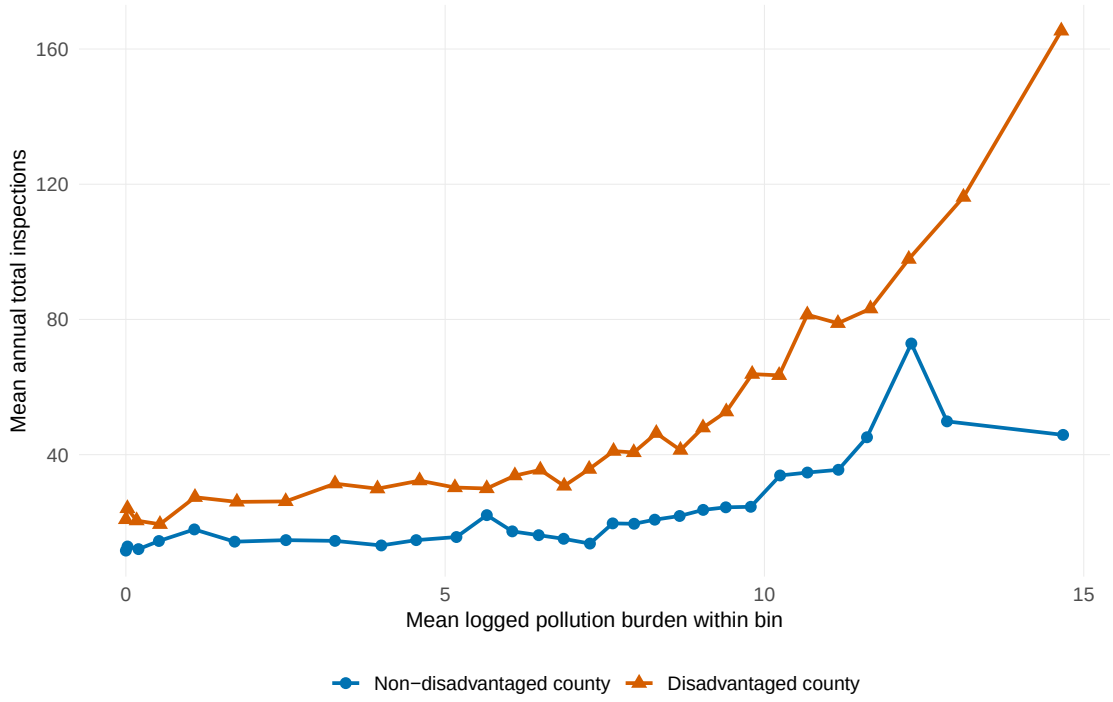


Figure B4: **Binned Relationship Between Pollution Burden and Inspections.** Each point reports the mean annual inspection count within one of 30 quantile bins of logged pollution burden, calculated separately for disadvantaged and non-disadvantaged county-year observations. The figure is descriptive and does not adjust for facility concentration, state fixed effects, or year fixed effects.

Appendix C. Additional Model Results

C.1 Full Negative Binomial Results

Table C1 reports sequential negative binomial specifications for annual total inspections. The models introduce facility concentration, state fixed effects, and the interaction between disadvantaged status and pollution burden in stages, providing a comparison of how the estimated disadvantage relationship changes as the specification becomes more fully adjusted.

Table C1: Sequential Negative Binomial Specifications

Term	(1)	(2)	(3)	(4)
Disadvantaged county	1.936*** [1.692, 2.215]	1.703*** [1.506, 1.927]	1.102* [1.002, 1.211]	1.107* [1.007, 1.217]
Logged pollution burden	1.131*** [1.115, 1.147]	1.025** [1.009, 1.042]	1.010* [1.000, 1.020]	1.007 [0.986, 1.028]
Disadvantaged × pollution burden				1.004 [0.983, 1.025]
Logged facilities		1.612*** [1.509, 1.722]	1.658*** [1.586, 1.734]	1.657*** [1.585, 1.733]
Observations	23,155	23,155	23,155	23,155
Counties	2,500	2,500	2,500	2,500
Year fixed effects	Yes	Yes	Yes	Yes
State fixed effects	No	No	Yes	Yes
County-clustered SEs	Yes	Yes	Yes	Yes

Notes: Entries are incidence-rate ratios, with county-clustered 95 percent confidence intervals reported in brackets beneath the estimates. Model (1) includes logged pollution burden and year fixed effects. Model (2) adds logged facility concentration. Model (3) adds state fixed effects. Model (4) adds the disadvantage-by-pollution-burden interaction. Blank cells indicate that the corresponding variable is not included in that specification. $+p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

C.2 Air Inspection Model

Table C2 reports the negative binomial estimates for annual air inspections using the same specification as the primary total-inspection model. At the average level of logged pollution burden, disadvantaged counties receive approximately 24.7 percent more air inspections than non-disadvantaged counties. The disadvantage-by-pollution-burden interaction remains statistically indistinguishable from one, providing no evidence that air inspection activity responds more strongly to increasing pollution burden in disadvantaged counties.

Table C2: Air Inspection Model

Term	Air inspections
Disadvantaged county	1.247** [1.058, 1.469]
Logged pollution burden	1.015 [0.965, 1.067]
Disadvantaged × pollution burden	1.009 [0.959, 1.062]
Logged facilities	1.945*** [1.836, 2.060]
Observations	23,155
Counties	2,500
Year fixed effects	Yes
State fixed effects	Yes
County-clustered SEs	Yes
Negative binomial theta	1.172

Notes: Entries are incidence-rate ratios, with county-clustered 95 percent confidence intervals reported in brackets beneath the estimates. Logged pollution burden is centered at its sample mean. The model includes logged facility count, state fixed effects, and year fixed effects. $+p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

C.3 Water Inspection Model

Table C3 reports the negative binomial estimates for annual water inspections using the same specification applied to total and air inspection activity. At the average level of logged pollution burden, disadvantaged counties do not receive statistically distinguishable levels of water inspections relative to non-disadvantaged counties (IRR = 1.022, 95% CI [0.929, 1.124]). The disadvantage-by-pollution-burden interaction is effectively one (IRR = 1.000, 95% CI [0.979, 1.021]), providing no evidence that water inspection activity becomes more

responsive to pollution burden in disadvantaged counties. Facility concentration remains strongly associated with water inspection activity, with an IRR of 1.503.

Table C3: Water Inspection Model

Term	Water inspections
Disadvantaged county	1.022 [0.929, 1.124]
Logged pollution burden	1.005 [0.985, 1.026]
Disadvantaged × pollution burden	1.000 [0.979, 1.021]
Logged facilities	1.503*** [1.430, 1.579]
Observations	23,155
Counties	2,500
Year fixed effects	Yes
State fixed effects	Yes
County-clustered SEs	Yes
Negative binomial theta	1.538

Notes: Entries are incidence-rate ratios, with county-clustered 95 percent confidence intervals reported in brackets beneath the estimates. Logged pollution burden is centered at its sample mean. The model includes logged facility count, state fixed effects, and year fixed effects. $+p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

Appendix D. Panel and Sample Robustness

D.1 Balanced-Panel Results

Table D1 reports the primary negative binomial specification using only counties observed in every year from 2013 through 2022. The balanced sample contains 2,150 counties and

21,500 county-year observations. The estimated incidence-rate ratio for disadvantaged counties is 1.113 (95% CI [1.007, 1.230]), closely matching the primary-sample estimate. The disadvantage-by-pollution-burden interaction also remains statistically indistinguishable from one. These results indicate that incomplete county-year coverage in the primary sample does not materially account for the estimated relationship between disadvantage and inspection activity.

Table D1: Balanced-Panel Model

Term	Balanced panel
Disadvantaged county	1.113* [1.007, 1.230]
Logged pollution burden	1.016 [0.993, 1.039]
Disadvantaged × pollution burden	0.998 [0.975, 1.022]
Logged facilities	1.658*** [1.583, 1.737]
Observations	21,500
Counties	2,150
Year fixed effects	Yes
State fixed effects	Yes
County-clustered SEs	Yes
Negative binomial theta	2.109

Notes: Entries are incidence-rate ratios, with county-clustered 95 percent confidence intervals reported in brackets beneath the estimates. The balanced sample contains counties observed in every year from 2013 through 2022. Logged pollution burden is centered using the mean from the primary analytic sample.

D.2 Full-Sample and Geographic-Sample Comparisons

Table D2 compares the primary specification with a model retaining all available geographic units, including the U.S. territories excluded from the main analysis. The estimated association between disadvantaged status and inspection activity is nearly identical across the two samples. The incidence-rate ratio for disadvantaged counties is 1.107 in the primary sample and 1.105 in the full geographic sample, while the disadvantage-by-pollution-burden interaction remains close to one in both specifications. The geographic restrictions applied to the primary sample therefore do not materially affect the estimated relationship between disadvantage and regulatory inspection activity.

Table D2: Full-Sample and Geographic-Sample Comparisons

Term	Primary sample	All geographies
Disadvantaged county	1.107* [1.007, 1.217]	1.105* [1.005, 1.215]
Logged pollution burden	1.007 [0.986, 1.028]	1.006 [0.986, 1.027]
Disadvantaged × pollution burden	1.004 [0.983, 1.025]	1.003 [0.982, 1.024]
Logged facilities	1.657*** [1.585, 1.733]	1.663*** [1.591, 1.738]
Observations	23,155	23,631
Counties	2,500	2,551
Geographic coverage	50 states + D.C.	States + D.C. + territories
Year fixed effects	Yes	Yes
State-jurisdiction fixed effects	Yes	Yes
County-clustered SEs	Yes	Yes

Notes: Entries are incidence-rate ratios, with county-clustered 95 percent confidence intervals reported in brackets beneath the estimates. The primary sample contains counties and county-equivalent units in the 50 states and the District of Columbia. The all-geographies specification additionally retains the U.S. territories available in the merged dataset. $+p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

D.3 Extreme-Value Sensitivity

The distribution of both pollution burden and inspection activity contains a small number of unusually large county-year observations. Table D3 re-estimates the primary specification after excluding observations above the ninety-ninth percentile of pollution burden and total inspections. The estimated incidence-rate ratio for disadvantaged counties is 1.104 (95% CI [1.010, 1.207]), closely matching the primary-sample estimate of 1.107. The disadvantage-by-pollution-burden interaction also remains effectively one. The main relationship between

disadvantage and inspection activity is therefore not driven by the most extreme observations in the pollution-burden or inspection distributions.

Table D3: Extreme-Value Sensitivity

Term	Trimmed sample
Disadvantaged county	1.104* [1.010, 1.207]
Logged pollution burden	1.010 [0.990, 1.030]
Disadvantaged × pollution burden	1.001 [0.981, 1.022]
Logged facilities	1.602*** [1.543, 1.664]
Observations	22,735
Counties	2,489
Sample restriction	Exclude observations above 99th percentile
Year fixed effects	Yes
State fixed effects	Yes
County-clustered SEs	Yes
Negative binomial theta	2.130

Notes: Entries are incidence-rate ratios, with county-clustered 95 percent confidence intervals reported in brackets beneath the estimates. The sensitivity sample excludes county-year observations above the ninety-ninth percentile of pollution burden and total inspections. The specification otherwise matches the primary model and includes logged facility count, state fixed effects, and year fixed effects. $+p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

D.4 Continuous Disadvantage Measure

The primary specification uses CEJST's binary county classification, which equals one when a county contains at least one census tract designated as disadvantaged and zero otherwise.

This measure captures the presence of disadvantage within a county but does not distinguish counties containing a single disadvantaged tract from counties in which most or all tracts are disadvantaged. As a robustness check, the primary model is re-estimated using a continuous measure equal to the share of a county's census tracts identified as disadvantaged, constructed from the CEJST 2.0 tract-level release and aggregated to the county level using county FIPS codes. Because counties with few census tracts can reach a high disadvantaged-tract share with only a small number of qualifying tracts, this specification also includes the logged number of census tracts as a covariate to separate disadvantage concentration from county size.

Table D4 reports the results. The estimated association between disadvantage concentration and total inspections is not statistically distinguishable from zero (IRR = 0.898, 95% CI [0.797, 1.012]). The interaction between disadvantage concentration and pollution burden is negative and statistically significant (IRR = 0.976, 95% CI [0.956, 0.996]), indicating that inspection activity becomes somewhat less responsive to rising pollution burden in counties with a greater concentration of disadvantaged tracts. This result qualifies the primary finding reported in the main text (Table 2). While the binary disadvantage classification is associated with modestly higher baseline inspection levels, the continuous measure indicates that more concentrated disadvantage is associated with weaker regulatory responsiveness to environmental burden, a pattern the binary measure does not detect. The two measures therefore appear to capture different dimensions of the disadvantage-inspection relationship, baseline attention and burden responsiveness, rather than producing directly contradictory findings.

Table D4: Robustness Using a Continuous Disadvantage Measure

Term	Continuous CEJST measure
Disadvantaged tract share	0.898+ [0.797, 1.012]
Logged pollution burden	1.007 [0.998, 1.015]
Disadvantaged tract share × pollution burden	0.976* [0.956, 0.996]
Logged tract count	1.443*** [1.383, 1.506]
Logged facilities	1.252*** [1.196, 1.311]
Observations	23,155
Counties	2,500
Year fixed effects	Yes
State fixed effects	Yes
County-clustered SEs	Yes
Negative binomial theta	2.261

Notes: Entries are incidence-rate ratios, with county-clustered 95 percent confidence intervals reported in brackets beneath the estimates. Disadvantaged tract share is the proportion of a county's census tracts identified as disadvantaged by CEJST, constructed from the CEJST 2.0 tract-level release and aggregated to the county level, and is centered at its sample mean. Logged tract count is included to separate disadvantage concentration from county size. Logged pollution burden is centered at its sample mean. $+p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

D.5 Subperiod Stability of the Disadvantage Estimate

The CEJST disadvantage classification was released in 2022 and is applied as time-invariant across the 2013–2022 panel, as described in the main text (Section 3.5, Disadvantaged

County Classification). To assess whether this introduces temporal bias, the primary model is estimated separately for an early subperiod (2013–2017) and a late subperiod (2018–2022). Table D5 reports the results. The estimated incidence-rate ratio for disadvantaged counties is 1.106 (95% CI [1.002, 1.220]) in the early subperiod and 1.103 (95% CI [0.998, 1.218]) in the late subperiod, closely matching the primary-sample estimate of 1.107 reported in the main text (Table 2). The disadvantage-by-pollution-burden interaction remains statistically indistinguishable from one in both subperiods. The near-identical estimates across the two halves of the panel indicate that the reported association between disadvantage and inspection activity is not an artifact of applying a classification developed near the end of the study period retroactively to earlier years.

Table D5: Subperiod Stability of the Disadvantage Estimate

Term	2013-2017	2018-2022
Disadvantaged county	1.106* [1.002, 1.220]	1.103+ [0.998, 1.218]
Logged pollution burden	1.001 [0.979, 1.025]	1.009 [0.988, 1.031]
Disadvantaged × pollution burden	1.011 [0.988, 1.035]	0.999 [0.977, 1.022]
Logged facilities	1.685*** [1.612, 1.762]	1.636*** [1.558, 1.718]
Observations	11,575	11,580
Counties	2,413	2,424
Year fixed effects	Yes	Yes
State fixed effects	Yes	Yes
County-clustered SEs	Yes	Yes

Notes: Entries are incidence-rate ratios, with county-clustered 95 percent confidence intervals reported in brackets beneath the estimates. The sample is split into an early subperiod (2013-2017) and a late subperiod (2018-2022) to assess whether the time-invariant application of the CEJST disadvantage classification, released in 2022, introduces temporal bias. Logged pollution burden is centered at its sample mean. Both models include logged facility count, state fixed effects, and year fixed effects. $+p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

D.6 States with Within-State Variation

The CEJST county-aggregation rule produces substantial variation in the share of disadvantaged counties across states. Eighteen states have every county classified as disadvantaged, meaning those states contribute no within-state contrast for the state-fixed-effects specification to draw on, and thirty-three states have a disadvantaged-county share above 90 percent. Table D6 reports the primary model restricted to the thirty-three states containing at least one non-disadvantaged county.The estimated incidence-rate ratio for disadvantaged

counties is 1.125 (95% CI [1.024, 1.237]), closely matching and marginally exceeding the primary-sample estimate of 1.107 reported in the main text (Table 2). This indicates that the main finding is not an artifact of states with uniform disadvantage classification and is instead identified by genuine within-state variation in county disadvantage status.

Table D6: Robustness Restricted to States with Within-State Variation

Term	States with variation
Disadvantaged county	1.125* [1.024, 1.237]
Logged pollution burden	1.011 [0.990, 1.032]
Disadvantaged × pollution burden	0.994 [0.973, 1.016]
Logged facilities	1.595*** [1.506, 1.689]
Observations	17,142
Counties	1,868
States excluded	18
Year fixed effects	Yes
State fixed effects	Yes
County-clustered SEs	Yes

Notes: Entries are incidence-rate ratios, with county-clustered 95 percent confidence intervals reported in brackets beneath the estimates. The sample excludes eighteen states in which every county is classified as disadvantaged under CEJST, leaving no within-state contrast for the state-fixed-effects specification to draw on in those states. Logged pollution burden is centered at its sample mean. The model includes logged facility count, state fixed effects, and year fixed effects. $+p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

D.7 Robustness Controlling for County Poverty Rate

The theoretical argument in Section 2 emphasizes political and economic resources as a mechanism linking disadvantage to regulatory attention, but the primary specification does not include a direct measure of economic hardship. Table D7 reports the primary model with the addition of county poverty rate, measured as the mean percent of individuals below 200 percent of the Federal Poverty Line across a county's census tracts and constructed from the CEJST 2.0 tract-level release. Controlling for poverty rate strengthens rather than weakens the disadvantage estimate: the incidence-rate ratio for disadvantaged counties rises to 1.215 (95% CI [1.095, 1.349]) from the primary-sample estimate of 1.107 reported in the main text (Table 2). Poverty rate itself is negatively and significantly associated with inspection activity (IRR = 0.991 per percentage point, 95% CI [0.986, 0.996]), indicating that counties with higher poverty rates receive somewhat fewer inspections after accounting for disadvantage status, pollution burden, facility concentration, state effects, and year effects. This pattern correlates with the continuous CEJST measure reported in Appendix D.4 where measures of disadvantage concentration or economic hardship are associated with fewer inspections, while the binary CEJST classification is associated with more, suggesting the two dimensions of disadvantage captured by these measures relate to inspection activity in different ways.

Table D7: Robustness Controlling for County Poverty Rate

Term	Poverty-controlled
Disadvantaged county	1.215*** [1.095, 1.349]
Logged pollution burden	1.006 [0.985, 1.027]
Disadvantaged × pollution burden	1.003 [0.982, 1.025]
Logged facilities	1.634*** [1.561, 1.711]
County poverty rate (percent)	0.991*** [0.986, 0.996]
Observations	23,155
Counties	2,500
Year fixed effects	Yes
State fixed effects	Yes
County-clustered SEs	Yes

Notes: Entries are incidence-rate ratios, with county-clustered 95 percent confidence intervals reported in brackets beneath the estimates. County poverty rate is the mean percent of individuals below 200 percent of the Federal Poverty Line across a county's census tracts, constructed from the CEJST 2.0 tract-level release and aggregated to the county level. Logged pollution burden is centered at its sample mean. The model includes logged facility count, state fixed effects, and year fixed effects. $+p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

Appendix E. Alternative Facility Specification

E.1 Facility-Offset Model

The primary specification estimates the relationship between facility concentration and inspection activity rather than imposing proportionality between the two. Table E1 reports

an alternative specification that instead includes logged facility count as an exposure offset, constraining a proportional relationship between the number of regulated facilities and expected inspections.

Under the offset specification, the estimated incidence-rate ratio for disadvantaged counties is 0.944 (95% CI [0.847, 1.053]) and is statistically indistinguishable from one. The estimate therefore differs from the positive disadvantage relationship in the primary county-level model. Logged pollution burden is negatively associated with inspection activity in the offset specification (IRR = 0.947, 95% CI [0.924, 0.969]), while the disadvantage-by-pollution-burden interaction remains close to one (IRR = 0.975, 95% CI [0.951, 1.000]). These results indicate that the higher total inspection activity observed in disadvantaged counties does not translate into higher inspection activity relative to the number of regulated facilities when proportionality between facilities and expected inspections is imposed.

Table E1: Facility-Offset Model

Term	Facility offset
Disadvantaged county	0.944 [0.847, 1.053]
Logged pollution burden	0.947*** [0.924, 0.969]
Disadvantaged × pollution burden	0.975+ [0.951, 1.000]
Observations	23,155
Counties	2,500
Facility treatment	Offset
Year fixed effects	Yes
State fixed effects	Yes
County-clustered SEs	Yes
Negative binomial theta	1.641

Notes: Entries are incidence-rate ratios, with county-clustered 95 percent confidence intervals reported in brackets beneath the estimates. The specification includes log(facilities) as an offset, imposing a coefficient of one on facility exposure. Year and state fixed effects are included. $+p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

Appendix F. Count-Model Diagnostics

F.1 Overdispersion and Model Fit

Table F1 reports diagnostic statistics for the annual total-inspection outcome and compares the fit of the negative binomial and Poisson specifications. The inspection outcome exhibits substantial overdispersion, with its variance considerably exceeding its mean. The negative binomial model accommodates this additional dispersion and provides the preferred

specification used throughout the analysis.

Table F1: Count-Model Diagnostics

Statistic	Value
Mean total inspections	47.0049
Variance of total inspections	7019.2048
Variance-to-mean ratio	149.3293
Zero inspection observations	302.0000
Zero inspection share	0.0130
Negative binomial theta	2.0205
Negative binomial AIC	199886.7234
Poisson AIC	628503.1842
Poisson Pearson dispersion	31.3638

Notes: The variance-to-mean ratio and Poisson Pearson dispersion summarize overdispersion in annual inspection counts. The negative binomial specification permits the conditional variance to exceed the conditional mean.

Appendix G. Program-Specific Predicted Inspections

G.1 Predicted Air Inspections

Figure G1 presents average adjusted predictions from the air-inspection model across the observed pollution-burden distribution. Predicted air inspection activity is higher in disadvantaged counties throughout most of the distribution, consistent with the positive and statistically significant disadvantage coefficient reported in Table C2. The prediction curves also remain approximately parallel, reflecting the statistically insignificant disadvantage-by-pollution-burden interaction.

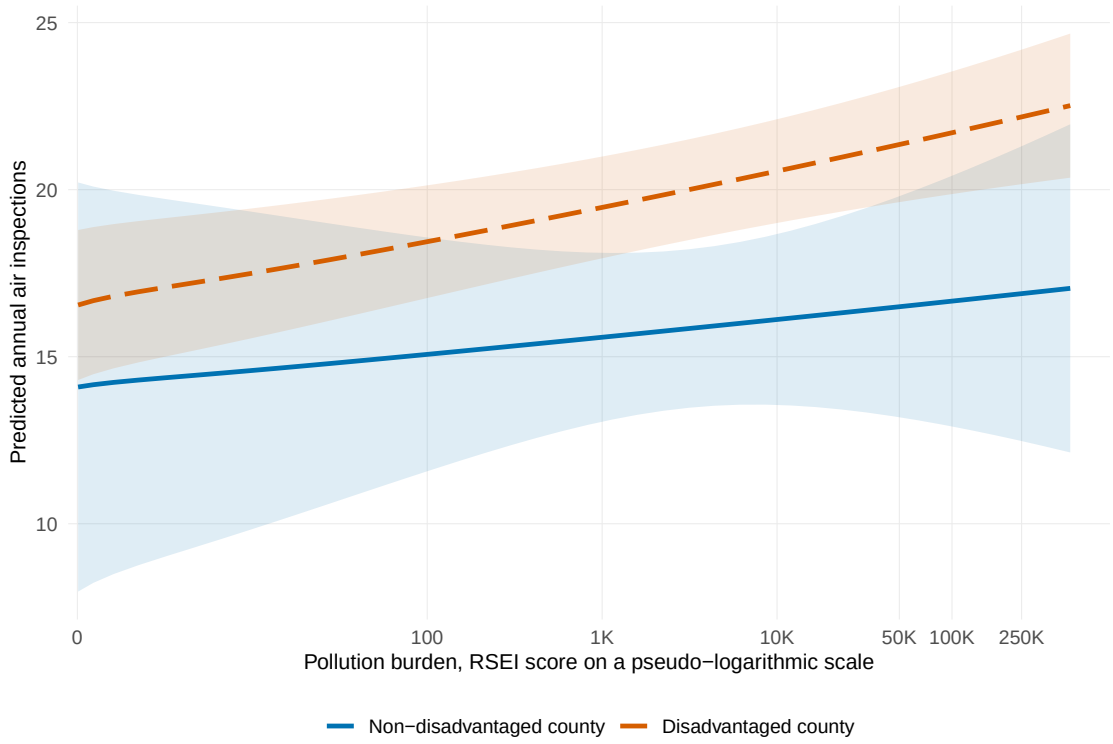


Figure G1: **Predicted Air Inspections Across Pollution Burden.** The figure presents average adjusted predictions from the air-inspection negative binomial model. Shaded bands report county-clustered 95 percent confidence intervals.

G.2 Predicted Water Inspections

Figure G2 presents average adjusted predictions from the water-inspection model across the observed pollution-burden distribution. Predicted water inspection activity is similar across disadvantaged and non-disadvantaged counties throughout most of the distribution, consistent with the statistically insignificant disadvantage coefficient reported in Table C3. The prediction curves also remain approximately parallel, reflecting the near-zero disadvantage-by-pollution-burden interaction.

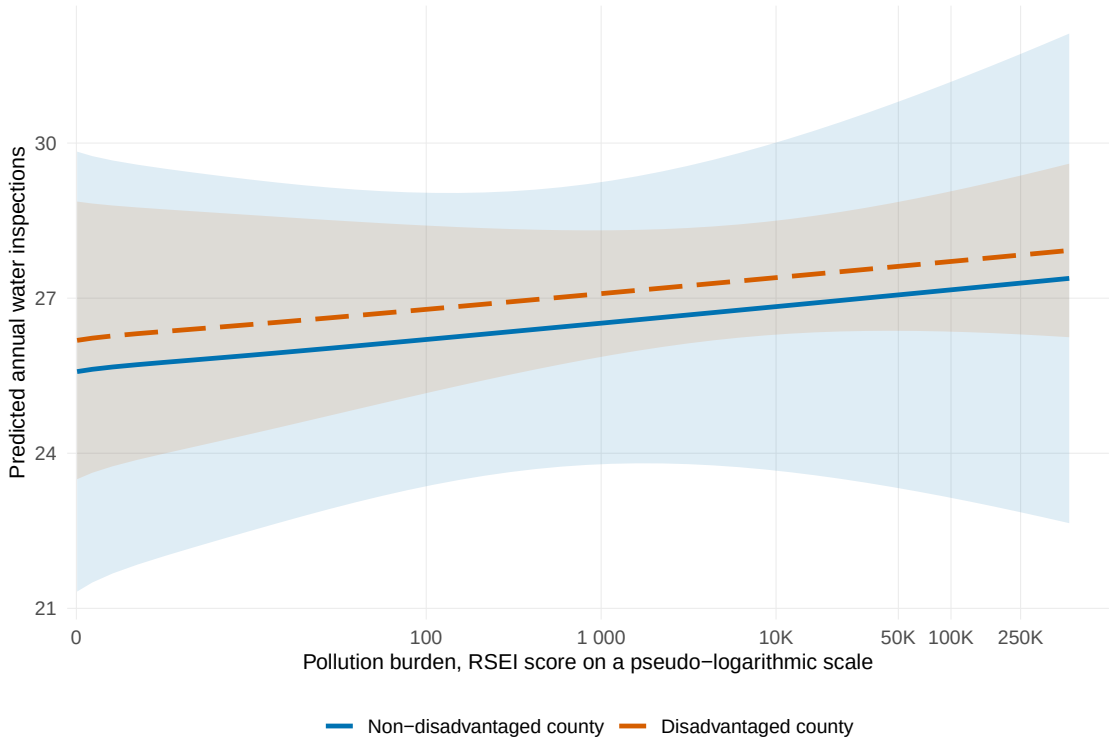


Figure G2: **Predicted Water Inspections Across Pollution Burden.** The figure presents average adjusted predictions from the water-inspection negative binomial model. Shaded bands report county-clustered 95 percent confidence intervals.

References

U.S. Environmental Protection Agency (2025a). *Enforcement and Compliance History Online (ECHO)*. Online database. URL: https://echo.epa.gov/.

U.S. Environmental Protection Agency (2025b). *Risk-Screening Environmental Indicators (RSEI) Model*. Online database and screening model. URL: https://www.epa.gov/rsei.

White House Council on Environmental Quality (2024). *Climate and Economic Justice Screening Tool: Technical Support Document*. Tech. rep. Version 2.0, December 2024. Washington, DC: White House Council on Environmental Quality. URL: https://climateprogramportal.org/wp-content/uploads/2025/02/cejst-technical-support-document.pdf.