Online Supporting Information
Gone with the Wind: Federalism and the Strategic Location of Air Polluters

James E. Monogan III
Assistant Professor
Univ. of Georgia
monogan@uga.edu

David M. Konisky
Associate Professor
Indiana Univ.
dkonisky@indiana.edu

Neal D. Woods
Associate Professor
Univ. of South Carolina
woodsn@mailbox.sc.edu

This supporting information reports several additional empirical results as well as more extensive information about our data. This document proceeds as follows:

Section 1 offers an illustration of a subset of our data by mapping the values of our dependent variable for four example states out of 48 studied.

Section 2 reports additional details about our primary model from the article, specifically describing our smoothed intercept term.

Section 3 reports an additional model based on our theory. We analyze the location of power plants sited before and after passage of the Clean Air Act of 1970. These results show evidence that the law seems to have diminished free riding among these facilities.

Section 4 reports inhomogenous Poisson models for the case observations and the control observations to disaggregate our finding somewhat. The count-oriented model of the cases serves to illustrate a model with no control group.

Section 5 reports the results for a model which considers that free riding might be especially heightened along a state border. The model in this section thereby allows a spline function of distance to affect the relative risk of an air polluter.

Section 6 reports the regression table for an interactive model with state-level covariates (visually presented in the article's Figure 3). It also reports models that interact state region and state wind speed with distance.

Section 7 reports the regression table for an additional model that adds two new predictors. Principally, this new model tests a hypothesis of whether population density affects probability of an air polluter, and also illustrates that the results are similar when fixed effects are included.

Section 8 reports spatial correlation measures—Moran's $\mathcal{I}$ and Geary's $\mathcal{C}$—for average residual by state, average level of free riding, and environmental policy.

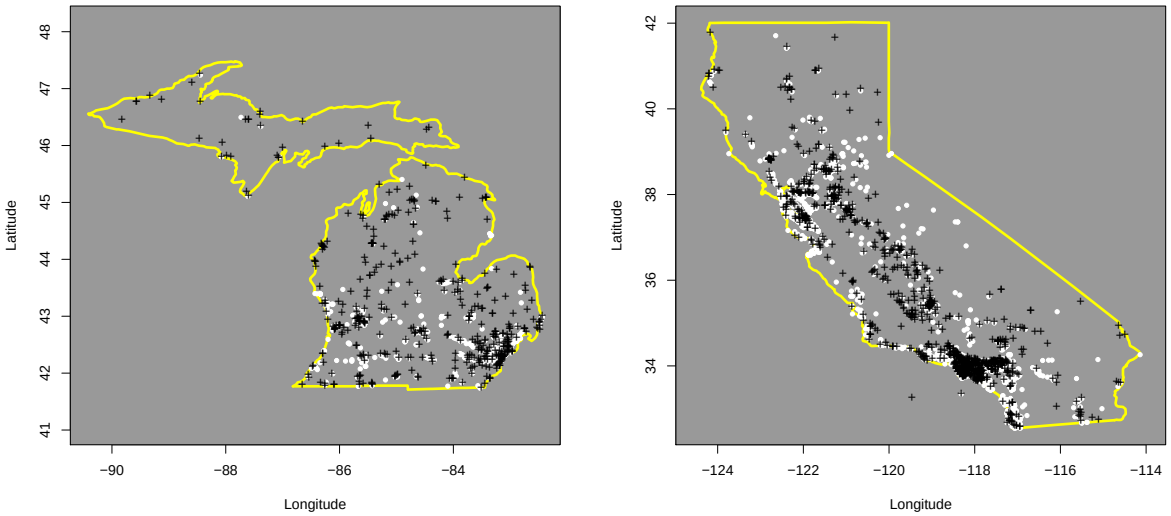
Section 9 offers details about the sources of data and the content of measures.

1 Maps of Example Data

We start by offering examples of how our dependent variable is structured. Figure A.1 presents the data in four selected states from the 48 we study: Michigan, California, Texas, and Indiana. Each panel shows a map that outlines the respective state's geographic borders in latitude and longitude. The black crosses represent the location of every major air polluter in the state, and the white dots represent every large quantity generator of hazardous waste. Starting in Subfigure A.1(a), the prevailing winds in Michigan tend to be eastbound. Observationally this figure shows that, wherever a concentration of control observations is located, major air polluters tend to be located to the east (or downwind) of these control observations. California also has generally eastward prevailing winds, and we see a similar pattern in Subfigure A.1(b). By contrast, most prevailing winds in Texas and Indiana have a northward direction. Hence, in Subfigures A.1(c) & A.1(d), we see that clusters of major air polluters tend to be situated closer to the northern border and a bit more northward than the nearby clusters of large quantity generators. Anecdotally, then, these visual patterns fit our theory that major air polluters tend to be farther downwind relative to other major industry. The point pattern models in the article offer a more systematic and rigorous test of this trend in general.

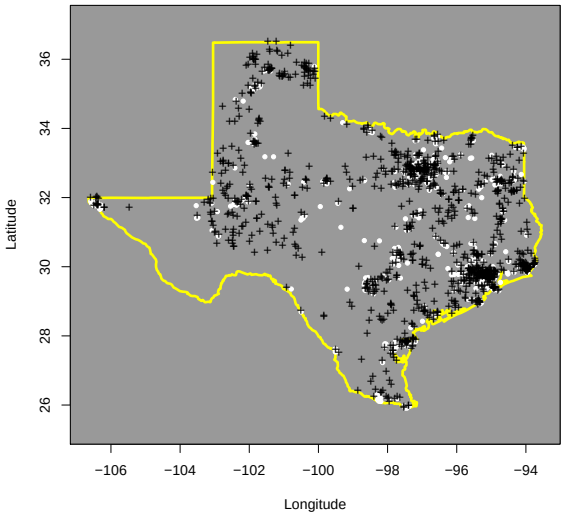
2 Additional Details from the Empirical Model

Figure A.2 presents additional information for the primary model in Table 1 from the main article, which uses scaled distance to the downwind border. This figure visually presents the nonparametric, smoothed generalized additive intercept that was included in the model. The horizontal axis of this figure presents longitude coordinates, and the vertical axis presents latitude coordinates. The map shows a point for every site of a major air polluter or hazardous waste generator in our data. The lighter-shaded areas indicate a higher baseline relative risk for hosting a major air polluter, and the darker-shaded areas indicated a lower baseline relative risk. The contour lines on the plot also serve to indicate depth, marking areas having the same value of the smoothed intercept. Again, higher values do not indicate more numerous major air polluters in raw counts, but that air polluters are likely to make up a higher proportion of the sites relative to hazardous waste facilities. As article Table 1 reports, this smoothed term easily explains a discernible amount of variance beyond the flat intercept model (as the significant χ^2 test indicates), and distance from the downwind border continues to have explanatory power even with the smoothed intercept in the model.

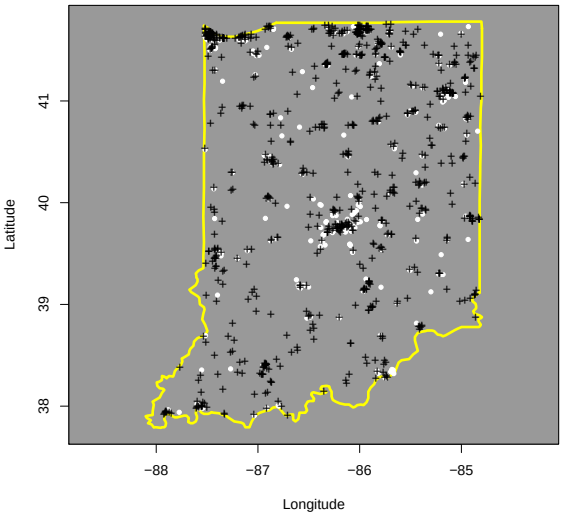


(a) Michigan

(b) California



(c) Texas



(d) Indiana

Figure A.1: Observed sites in four states. The outline in yellow draws the state, black crosses represent major air polluters, and white dots represent hazardous waste generators.

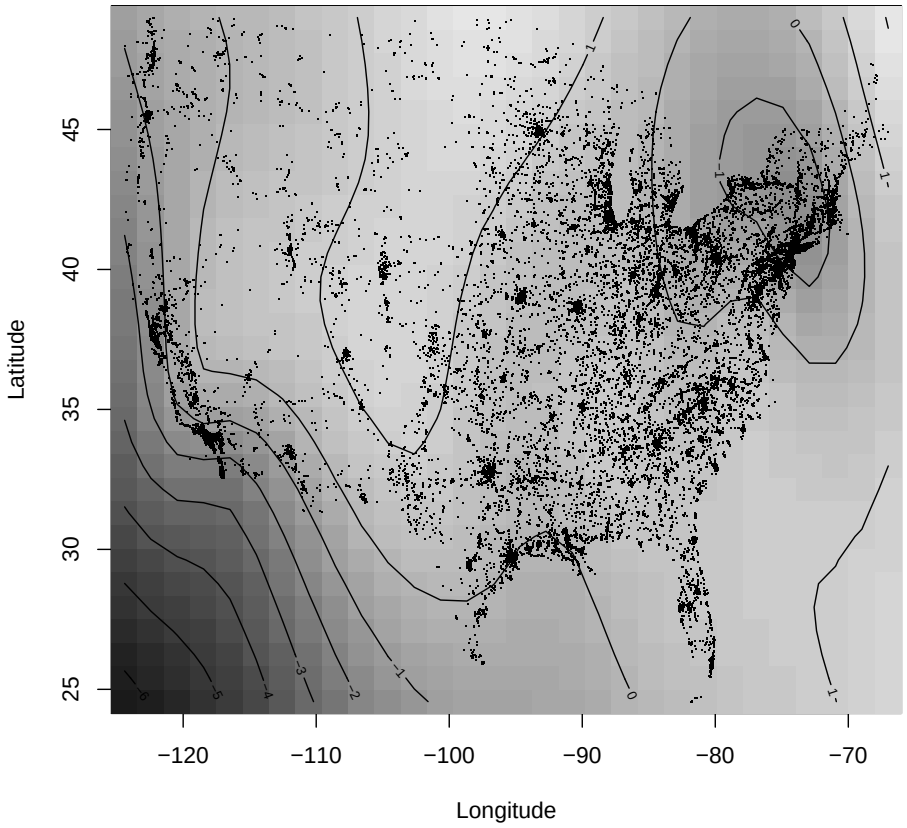


Figure A.2: Smoothed baseline risk surface for major air polluters in the United States, prior to accounting for distance. Lighter shading indicates a higher baseline propensity that a site will host a major air polluter. Points represent the location of observed sites.

3 Free Riding before Federal Policy Intervention

The logic of environmental federalism discussed in the article suggests that free riding should be more prevalent when greater environmental regulatory authority rests with subnational governments. Major U.S. environmental laws such as the Clean Air Act (CAA) allow the states a role in implementation and enforcement, subject to EPA oversight (Lowry 1992). Federal regulatory oversight has been found to vary in response to both state-level and national-level political and task factors (Hedge & Scicchitano 1994, Wood 1992), but in

general the presence of this oversight authority should reduce states' latitude to use environmental regulation as a means to induce locational free riding. Moreover, federal attention to the cross-border effects of air pollution should help alleviate locational free riding through other mechanisms as well. Thus we would expect to see less locational free riding after the passage of the federal Clean Air Act of 1970 than in the preceding years when there was minimal federal involvement in clean air policy.[1] Accordingly, we posit that facilities sited before the 1970 CAA should be located closer to downwind borders than facilities sited after this law.

To test whether environmental free riding was more pervasive before the federal CAA was enacted in 1970, we require information on not just where facilities are located, but when they came online. Although this information is not available for the full set of facilities we have been analyzing to this point, it is available from a separate data source of electric power plants. Specifically, the EPA's *Emissions & Generation Resource Integrated Database* contains detailed information on the year that power plants first went online.[2] We can use this information to analyze whether the pattern of plant location differed before versus after the passage of the CAA. These results cannot speak to the patterns across all industries, but can offer evidence of the impact of this statute on free riding in the electricity generation sector.

To evaluate the impact of time on evidence of free riding, we estimate three models. First, we subset our data to include only power plants constructed prior to 1970, along with all of our control observations. During this period before the CAA, we expect that incentives are particularly strong to free ride because pollution levels from each plant will be higher than afterward. Second, we estimate a model with the controls and the subset of power plants constructed during the decade following the adoption of the CAA to allow for an immediate post-policy rush and delayed implementation time. Third, we estimate a model with power plants constructed after 1980, along with all control observations.[3]

Table A.1 presents the results of these three models of power plant location, subsetted

[1] A principal reason Congress enacted the 1970 Clean Air Act was to address interstate air pollution, but devising regulations has proven to be a vexing issue for the EPA. Over the past decade, EPA regulatory efforts to address interstate air pollution have met stiff resistance from both industry and the federal courts. Recently, however, in the case of *Environmental Protection Agency v. EME Homer City Generation* (2014), the Supreme Court in a 6-2 decision upheld the EPA's 2011 Cross-State Air Pollution Rule, which was aimed at reducing power plant emissions that contribute to ozone and fine particulate pollution in other states. This reversed a Court of Appeals ruling that the EPA overstepped its authority with this rule, and ended more than a decade of federal court litigation on this issue.

[2] More specifically, the data provide the year in which each unit at a power plant went online. In this analysis, we use the year in which the first unit at each plant became operational. Source: http://www.epa.gov/cleanenergy/energy-resources/egrid/.

[3] We also considered a simpler design in which data were simply split between pre-1970 and post-1970. Under this subsetting specification, we still obtain the result of significant evidence of free riding before 1970 and no evidence of free riding after 1970. Theoretically, however, we believe allowing the ten year window is more plausible.

Table A.1: Models of Spatial Variation in Risk of Power Plant Location over Time (Generalized Additive Model, Logit Link)

	Model Specification		
Covariate	Before 1970	1970–1980	After 1980
Leeward distance	-0.3550^*	-0.3433^*	-0.0360
	(0.1057)	(0.2042)	(0.0920)
Intercept	-2.7665^*	-3.9838^*	-2.1983^*
	(0.0511)	(0.0917)	(0.0406)
AIC	10886	4166	15588
N_{total}	22445	21002	23434
N_{air}	1909	466	2898
$N_{hazardous}$	20536	20536	20536

Notes: Cell entries are coefficient estimates, with standard errors in parentheses. * means estimate is discernible at the 95% confidence level (one-tailed). Approximate significance test for baseline hazard smoothed over latitude and longitude is significant at the 95% confidence level in each model. Estimates computed with R 3.1.2.

by time periods. In all three of these models, the control group consists of all large quantity generators of hazardous waste from our entire sample because we do not have time-referenced data for control observations. In the first column, the case observations are power plants that went online prior to 1970. In the second column, the case observations are power plants that went online in the decade from from 1970 to 1980. In the third column, the case observations are power plants that went online after 1980. The predictor in every model is scaled distance to the leeward border, with the strongest effect being for pre-1970 plants. The effect was similar during the 1970s, and still negative and significant. With the plants that opened after 1980, though, the effect is small and not discernible from zero.

Figure A.3 presents the predicted probabilities that an industrial site will host a power plant based on these three separate analyses. The horizontal axis represents the scaled distance to the downwind border, and the vertical axis presents the probability of a power plant. The middle line shows the predicted probability of a power plant among plants constructed prior to 1970, with the shaded area representing the 90% confidence interval around the probability. This line shows the strongest effect, with upwind locations (closer to 1) being significantly less likely to host a power plant relative to downwind locations (closer to 0). The bottom line shows the predicted probability of a power plant when modeling the location of plants constructed during the 1970s. This effect is shallower than we see in the model of pre-1970s observations, though the effect is still negative and significant at the 95% one-tailed level. The top line shows the predicted probability of a power plant when modeling the location of plants constructed after 1980. By this time frame, there is

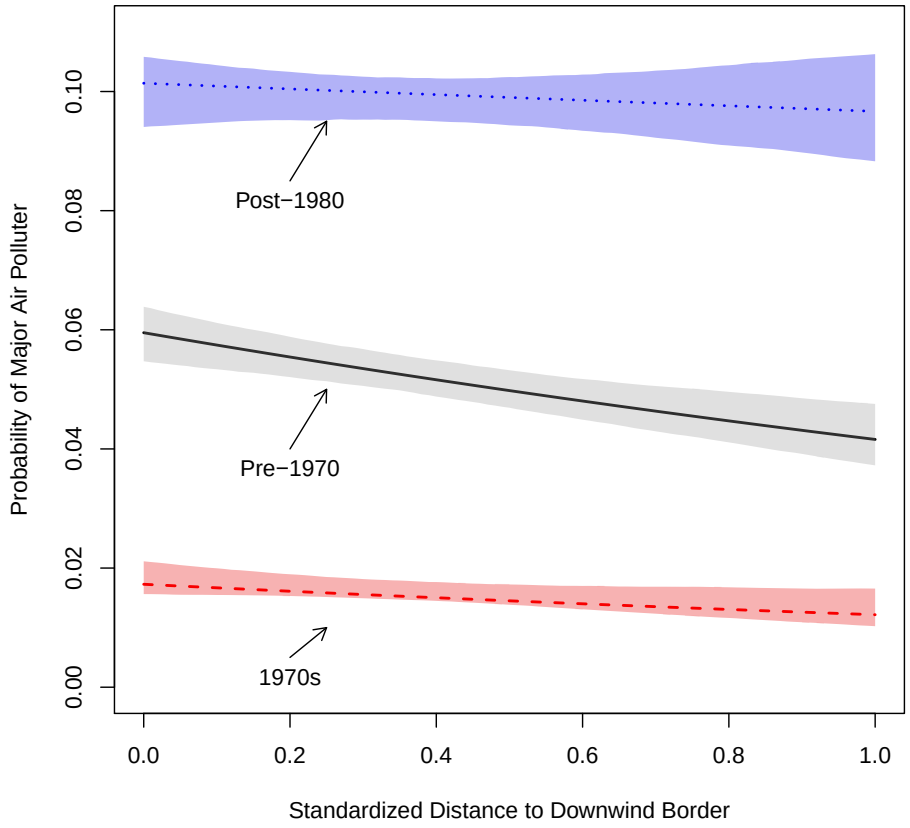


Figure A.3: Predicted probability that a power plant will emerge at an industrial site given scaled distance from downwind border. Predictions based on separate models that subset by the time the power plant was constructed. Shaded areas represent 90% confidence interval of forecast.

no discernible effect. All of this is consistent with the notion that the CAA helped curb the incentives to free ride. Free riding was prevalent in the energy industry before the CAA, diminished somewhat over the next ten years, and then became unobservable after that. Federal law, therefore, appears to have shaped site selection patterns in the energy sector, though we cannot speak to temporal patterns in other industries.

4 Count-Oriented Point Patterns of Cases & Controls

As an illustration of how an inhomogenous Poisson process can be estimated, we report the findings of such point pattern model in Table A.2. This count-oriented point pattern model does not include a control group, so it simply models the relative frequency with which major air polluters emerge in latitude and longitude. Like the models of relative risk presented in the paper, we include a smoothed nonparametric intercept in this model as well. Even with this data-driven term, we find a negative and discernible result for the coefficient on scaled distance from the leeward border. Hence, air polluters apparently become sparser the farther the distance from a state's downwind border. While this result fits with our hypothesis, the relative risk logit models presented in the paper are a better evaluation of free riding because the control group of large quantity generators offers an indication of how industry is distributed within a state. By evaluating how the distribution of air polluters differs from the reference group of other industrial sites we provide a fairer test of the hypothesis.

Table A.2: Model of Major Air Polluter Frequency, Inhomogenous Poisson Process with GAM Intercept

Covariate	Estimate	Std. Error	z-ratio	p-value
Scaled distance from leeward border	-0.3475	0.0721	-4.8180	0.0000
Intercept	0.8852	0.0494	17.9156	0.0000

Notes: N_{air}=16,211. AIC=-73570. Estimates computed with R 3.2.0.

By way of comparison, we also estimated an inhomogenous Poisson process model on our control group of large quantity generators of hazardous waste. This result is presented in Table A.3. As a useful contrast, notice that there is no discernible effect for scaled downwind distance in this model of our control group. Since the wind cannot carry away these sites' pollution, there is no incentive to push them leeward. Also, by way of illustration, if we computed $\lambda_1(\mathbf{x})$ and $\lambda_2(\mathbf{x})$, the spatially-varying intensity parameter for the case and control Poisson models, then recall that the ratio would give us the relative risk that we model with logistic regression in the main article. (Though such a procedure is more intensive computationally.) These models thereby disaggregate our final result into component parts.

Table A.3: Model of Large Quantity Generator Frequency, Inhomogenous Poisson Process with GAM Intercept

Covariate	Estimate	Std. Error	z-ratio	p-value
Scaled distance from leeward border	0.0015	0.0648	0.0232	0.9815
Intercept	-0.0987	0.0576	-1.7138	0.0866

Notes: $N_{hazardous}$=20,536. AIC=-127901. Estimates computed with R 3.2.0.

5 Threshold Model

In this section, we consider the notion that our results may be driven to a large degree by observations particularly close to the border. That is, within a certain threshold of the downwind border, the tendency for free riding might be heightened. To do this, we reconsidered our scaled distance from the downwind border measure, which ranges from 0 to 1. We estimated models of a spline function in which the knot point varied from various values from 0 to 0.5, figuring anything larger was no longer "close" to the border. The best-fitting model came with a knot point at 0.1. The results of this model are presented in Table A.4.

Table A.4: Model of Spatial Variation in Risk of Major Air Polluter Placement with Spline Effect for Scaled Distance (Generalized Additive Model, Logit Link)

Covariate	Coefficient	Std. Err.	z-ratio	p-value
Intercept	0.1138	0.0452	2.5199	0.0117
Scaled distance from leeward border	-4.3646	0.5321	-8.2023	0.0000
Switch in effect for scaled distance greater than 0.1	4.3244	0.5574	7.7586	0.0000

Notes: N_{total}=36,747 (N_{air}=16,211 air polluters and $N_{hazardous}$=20,536 hazardous waste large quantity generators). AIC=44714. Two-tailed p-values reported. Approximate significance test for baseline hazard smoothed over latitude and longitude is significant at the 95% confidence level ($\chi^2_{28.71} = 4771.45$). Estimates computed with R 3.2.0.

Functionally, this model assumes that distance to the downwind border had one effect until it hits the knot point (0.1 in this case), and then the linear effect abruptly changes. For sites located close to the border, no farther than one-tenth of the longest site distance in the state, the coefficient of -4.36 is the strongest effect we see in any of these results. Substantively, for states like Alabama, Louisiana, Montana, and Wisconsin, a standardized distance of 0.1 amounts to about 50 kilometers from the downwind border. This discernible effect shows that close to the border there is a particularly strong tendency to push air polluters downwind. For any distance from the border larger than one-tenth the longest distance (e.g., farther than 50 km in a state like Alabama), the effect nearly zeroes-out. By adding the base coefficient to the spline coefficient, we see that the coefficient drops to -0.04. While signed as we would expect, this effect is no longer discernible.[4] This model therefore implies that the sites closest to a state's downwind border are where the biggest free-riding action is.

As an illustration of how this works, Figure A.4 shows the functional form implied by the model estimated in Table A.4. 10 functional forms are illustrated because we considered

[4]The composite standard error of 0.05 gives us $z = -0.74$ and $p = 0.23$.

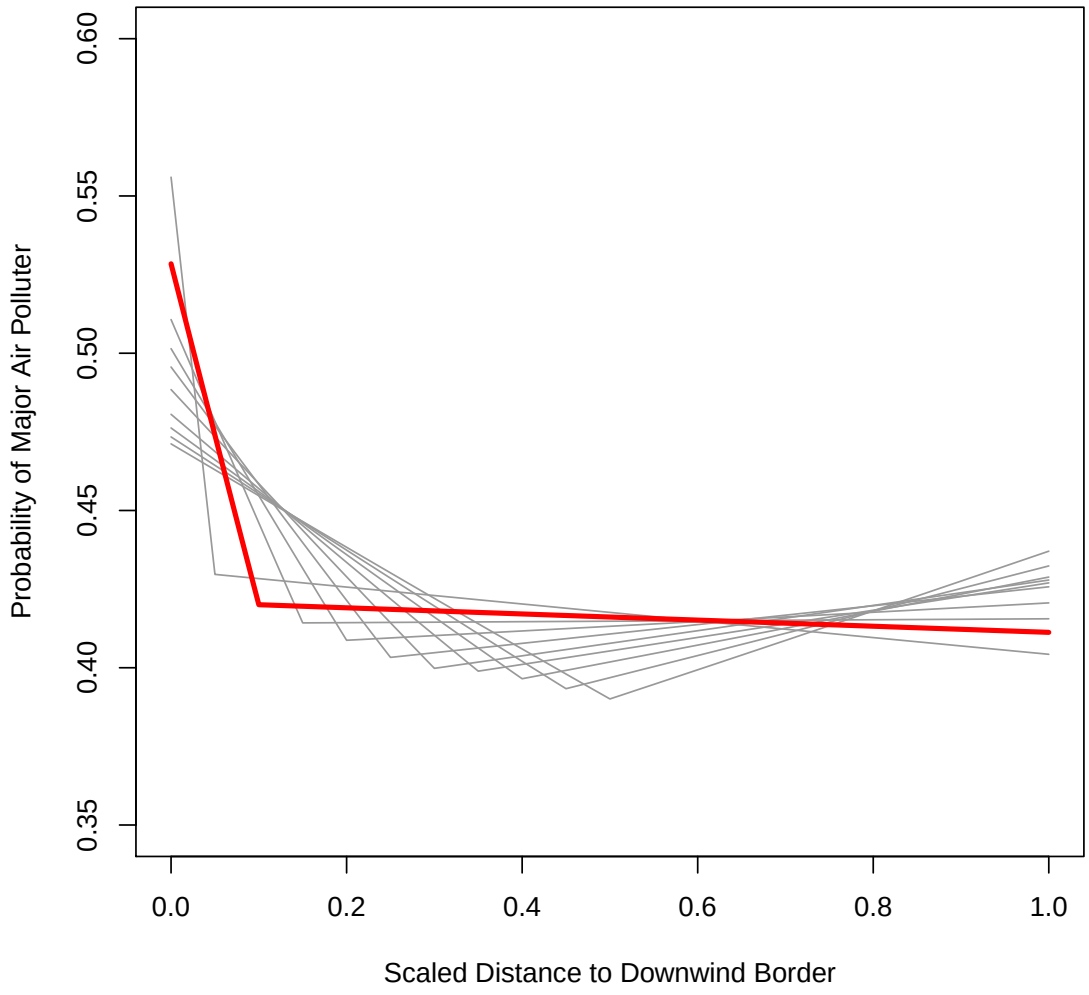


Figure A.4: Probability of a site hosting a major air polluter, as opposed to large quantity generator, assuming a spline effect of scaled downwind distance. Best fitting model (knot at 0.1) is represented with a bold red line. Alternative models are represented with light gray lines.

knots ranging from 0.05 to 0.5 in increments of 0.05. The results of each functional form are shown with a light gray line, except the best-fitting functional form, with a knot point of 0.1, is shown with a bold red line. Focusing on the form that best fit the data, in the

bold red line, from distances of 0 to 0.1, the probability of hosting a major air polluter drops considerably. Right on the border, the probability of hosting a major air polluter is 0.53, and by the one-tenth distance on the standard scale this probability has dropped to 0.42. From the one-tenth distance all the way out to the farthest distance from the border, the probability only drops to 0.41. Hence, the effect is considerably stronger in this area near the border and pretty flat afterward. This functional form also implies that near the border there is a heightened relative risk for air polluters relative to the rest of the entire state, which also fits our theory.

6 Interactive Models

Table A.5 presents the results of a model that interacts scaled downwind distance with three state-level variables: the Green Index of environmental policy, an index of locational economic development, and the number of environmental interests.[5] These results are interpreted in the article with the forest plot of Figure 3. The significant interaction terms indicate that free riding is stronger in states with less environmentally-friendly policy, states that pursue locational economic development strategies, and states with a heavy density of environmental lobbyists. For a state at the average level on these covariates, the coefficient for distance to downwind border becomes -0.2193, which is similar in size to our original result, and the effect remains discernible at any common confidence level. In fact, the post hoc z-test for this effect is: $z = \frac{-0.2193}{0.0524} = -4.1815$. The effect when all variables are held at their mean is visualized in all three panels of the article's Figure 3 with the conditional coefficients marked, "Mean."

It is worth reiterating that a state with a low score on the Green Index is not less likely to have air polluters overall, just more likely to reallocate them near borders. As Figure A.5 shows, on average states with a higher policy score on the green index tend to have a lower proportion of air polluters. Brown states that score low on this index, however, have a relatively high proportion of air polluters. With that relatively high proportion, then, it would make sense that these states would need to use any permitting or locational development strategies they have available to push as many of them downwind as possible—otherwise in-state air quality will suffer badly. Thus, given the relatively high proportion of air polluters, the strong free-riding effect for brown states implied by Table A.5 makes a lot of sense.

We also considered that attributes that vary regionally might alter the degree of free riding. To consider this, we fitted two more interactive models. One in which an indicator for the Census region of a state is interacted with distance to the downwind border, and the other in which wind velocity is considered to potentially alter the relationship. First,

[5]While the Green Index was measured as a snapshot in time, environmental policy is slow to change. Hence, states are stable in their overall level of green-friendly policy from year-to-year.

Table A.5: Model of Major Air Polluter Placement with Distance to Downwind Border, Interacted with State-Level Covariates

Covariate	Estimate	Std. Error	z-ratio	p-value
Scaled distance from leeward border	0.2111	0.4402	0.4795	0.6842
Green index	-0.0488	0.0060	-8.0731	0.0000
Locational development	-0.7109	0.1243	-5.7197	0.0000
Environmental interests	-0.0239	0.0038	-6.3123	0.0000
Distance×green	0.0148	0.0088	1.6775	0.0467
Distance×locational	-0.3432	0.2301	-1.4914	0.0679
Distance×interests	-0.0127	0.0060	-2.1110	0.0174
Intercept	2.2734	0.2391	9.5078	0.0000

Notes: Estimates computed with R 3.1.2. Approximate significance test for intercept smoothed over latitude and longitude: $\chi^2_{28.54} = 1999.084$ ($p < .001$). N_{total}=36,747 (N_{air}=16,211 air polluters and $N_{hazardous}$=20,536 hazardous waste large quantity generators). AIC=44482.

on region, we simply created four indicators for Census region and interacted them with standardized downwind distance. The implied coefficients for each region are reported in the forest plot of Figure A.6. As Figure A.6 shows, the strongest negative effect is in the Northeast, followed by the Midwest and then the South. The West actually has a positive coefficient, albeit an insignificant one. Therefore, we can see that the size of this effect varies with region.

We also consider the notion that average wind velocity can affect states' tendency to free ride, so we consider the state-level measure of average wind velocity.[6] It is difficult to get a good sense of how wind speed alone will affect free riding because several factors affect where pollution goes. Plume modeling normally factors in smokestack height, temperature, and topography as well, which we do not have measures for. Therefore, interpretation of our wind results should be tempered by the knowledge that atmospheric chemistry is complicated.

That said, Figure A.7 presents the results from a model that interacts the average wind speed in a state with distance to the downwind border.[7] The horizontal axis represents the observed values of average wind speed by state, ranging from the minimum of 10.5 km/h to a maximum of 26.3 km/h. The vertical axis represents the values of the coefficient for downwind distance in the logit model we estimated. With the solid black line we show the estimated effect of downwind distance given each level of average wind speed. With

[6]This is computed with the average wind speed of all NOAA-reporting weather stations within each state, using the data from the 299 stations recorded in *Climatic Wind Data for the United States* (NOAA, November 1998).

[7]We use the unstandardized measure of distance in kilometers because wind speed is measured based on actual distance in kilometers per hour, so the measure of downwind distance should also be computed using actual distance in kilometers.

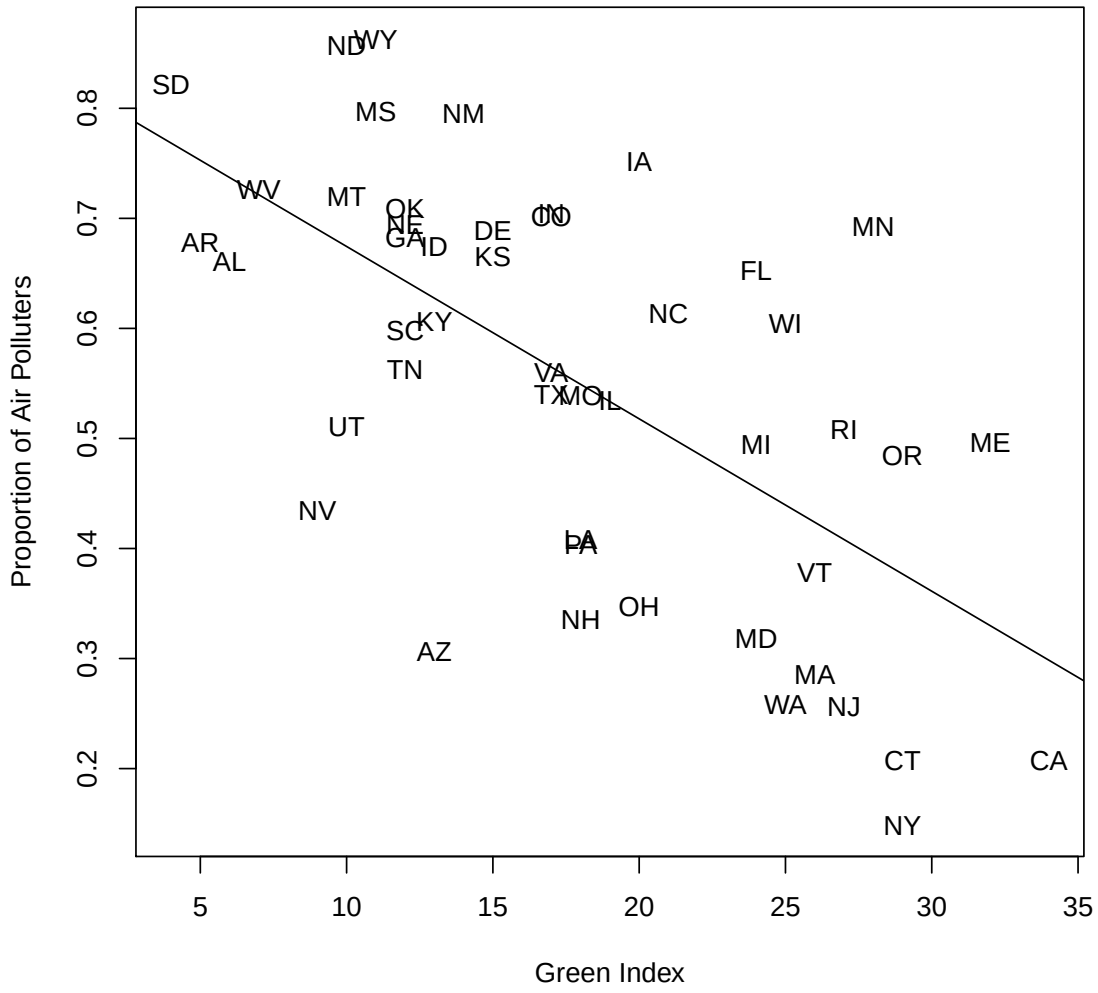


Figure A.5: Air polluters by state as a proportion of air and hazardous waste polluters, against the state's score on the green index.

the dashed red line we show the one-tailed 95% upper confidence bound of the estimated coefficient using the procedure described by Neyman (1937, 374-377).[8] Whenever the upper

[8]Neyman (1937) explains that by forming the upper bound of a 90% two-tailed confidence interval, we can say with 95% confidence that the coefficient could take on any value less than that. So imposing no lower limit makes the range from the upper bound to negative infinity a one-tailed 95% confidence region.

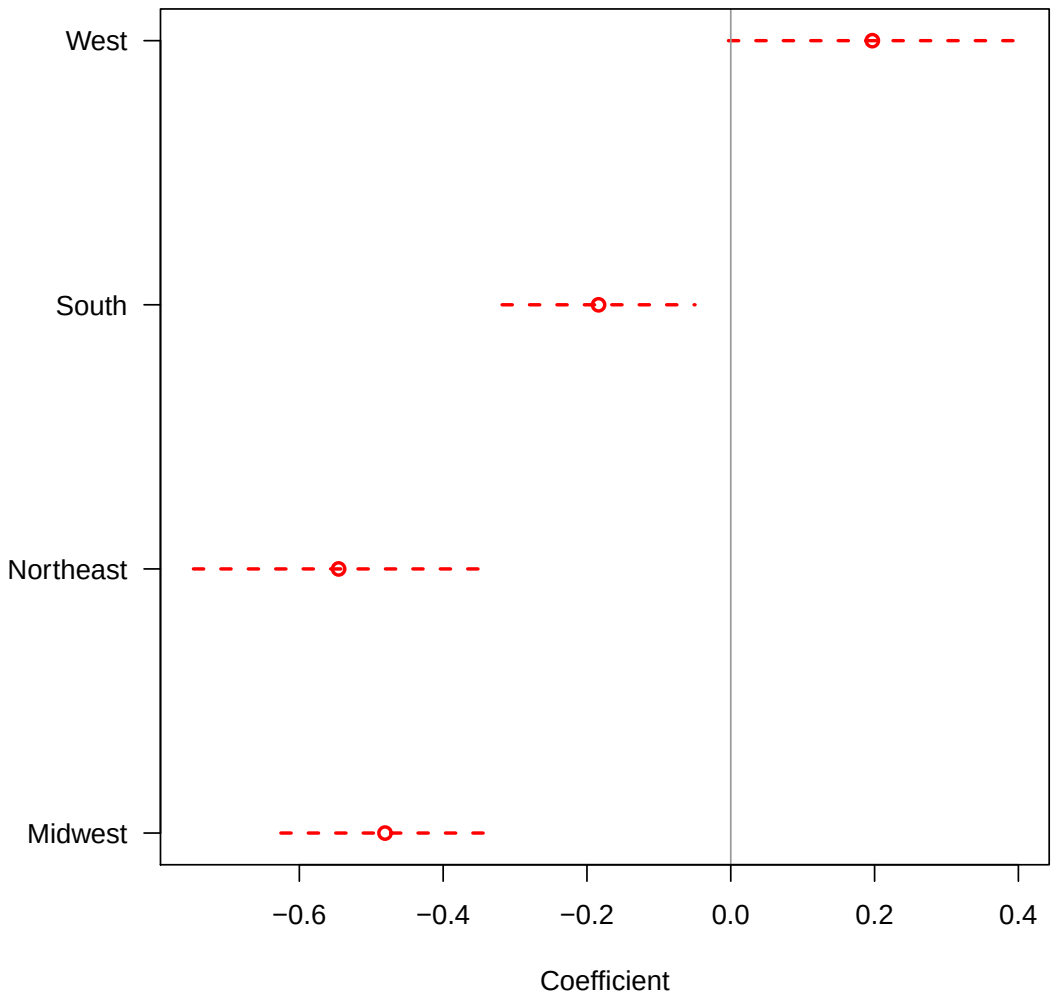


Figure A.6: Effect of standardized distance from downwind border contingent on Census region. Points represent the resulting coefficient for the effect of downwind distance in each region, and the dashed red line represents the two-tailed 90% confidence interval for the coefficient, which we use to conduct a one-tailed test for a negative effect at the 95% confidence level.

confidence bound is less than zero, we can say with 95% confidence that the effect is negative and significant.

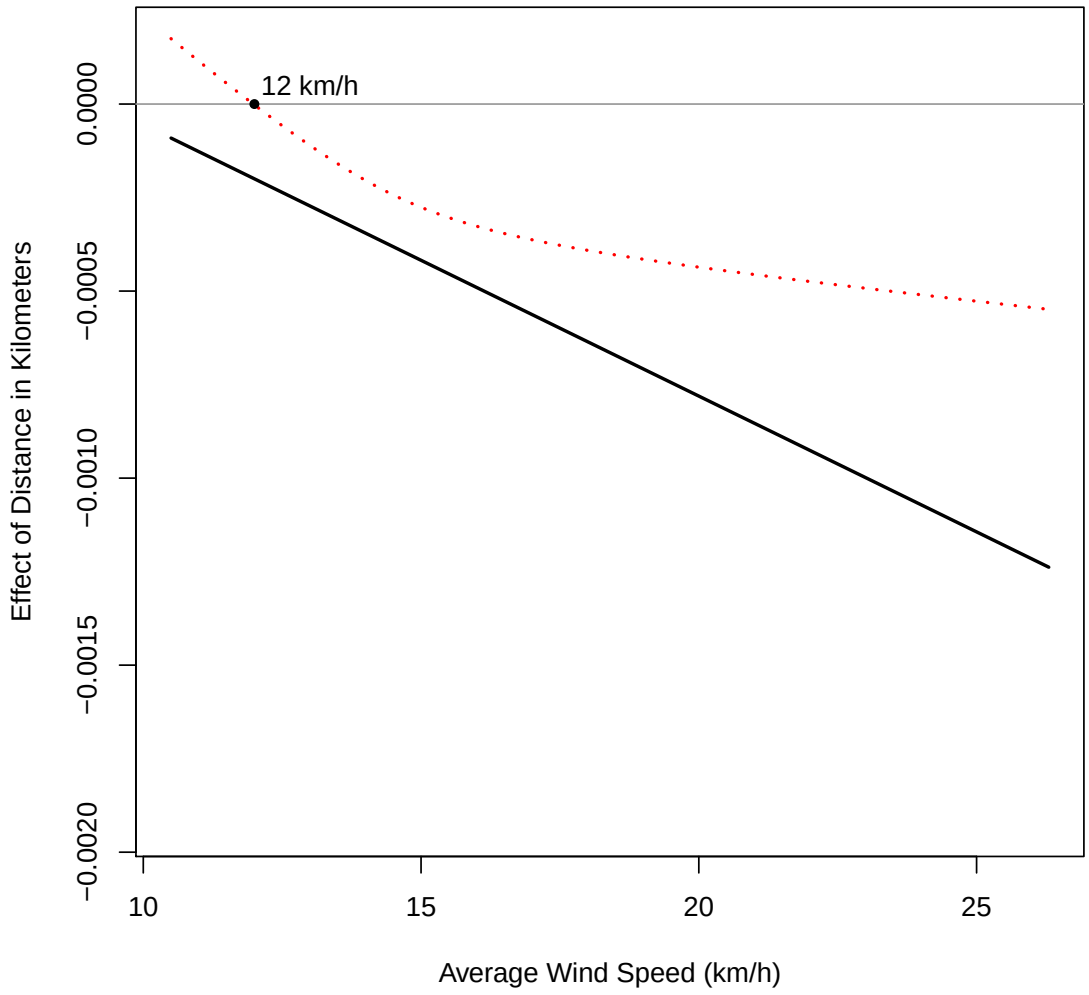


Figure A.7: Effect of kilometers from downwind border contingent on average wind speed in the state. The solid black line represents the resulting coefficient for the effect of downwind distance, and the dashed red line represents the one-tailed 95% upper confidence bound for the coefficient.

As Figure A.7 shows, as wind speed increases, the tendency to free ride strengthens. Specifically, a negative effect indicates a stronger tendency to free ride, and as wind speed increases the negative effect becomes more pronounced in this model. Absent other relevant

factors to plume modeling, this effect generally makes sense because stronger wind speeds are more able to quickly push air pollution away, thereby increasing the ability to free ride. Testing the hypothesis that the effect is negative, the 95% confidence bound is less than zero whenever the average wind speed is greater than 12 km/h. Thus, the free riding effect is not statistically discernible in states with still winds, but it is discernible in every state with average wind speeds greater than 12 km/h. The median wind speed in our data is 14.5 km/h, and 70% of the sites are in states with a higher wind speed than 12 km/h. So overwhelmingly we see discernible effects, and in the windiest states the effect is strongest.

7 Considering Population Density and State Effects

The model in Table A.6 revisits our primary model of interest from Table 1 in the main text, which used scaled distance from leeward border, and adds control variables. We would argue that control variables are not necessary for two reasons: First, we compare major air polluters to a similar control group of large quantity generators of hazardous waste, so that already holds most things equal. Second, the smoothed baseline hazard term in these models should pick up the impact of omitted covariates that would still be relevant. However, explaining more variance in the structural portion of the model based on theoretical reasons has the advantage of allowing us to test additional hypotheses of interest and reduces our reliance on a data-driven estimator. The first control is to include state-level fixed effects. The results of the coefficients on the 47 indicator variables (Alabama is the reference state) are omitted to preserve space. However, these indicators completely absorb any state-level predictor that might be relevant. The second control is the population density of the county a facility is sited in, measured in thousands of residents per square mile. As a county becomes more densely populated, holding state and location constant, industrial sites in the county become less likely to be hosts of major air polluters. If the population density increased by 1,000 residents per square mile, the odds of a major air polluter relative to a large quantity generator would drop by 3.8% on average and *ceteris paribus*. Meanwhile, our primary estimate of interest, the coefficient for scaled distance to the downwind border, is actually a bit stronger in this alternative specification. This model with controls implies that moving from a site on the state border to the site farthest upwind diminishes these odds by 23.7%, on average and *ceteris paribus*. Moving even a tenth of this distance diminishes the odds by 2.7%, on average and all else equal. Overall, this model offers further support for our results.[9]

[9]Despite the addition of these control variables, the smoothed baseline hazard term is still significant at the 95% confidence level ($\chi^2_{28.11} = 452.71$). It is also worth noting that the AIC indicates a better fit with these control variables. As a final point, we obtain similar results in models that use fixed effects alone or population density alone.

Table A.6: Model of Spatial Variation in Risk of Major Air Polluter Placement, Including Fixed Effects for States (Generalized Additive Model, Logit Link)

Covariate	Estimate	Std. Error	z-ratio	p-value
Intercept	-0.0148	0.2386	-0.0621	0.9505
Scaled distance from leeward border	-0.2707	0.0691	-3.9181	0.0001
Population density (1000/square mile)	-0.0388	0.0046	-8.4473	0.0000

Notes: Estimates computed with R 3.2.0. Model includes an indicator variable for every state except for Alabama, which is the reference group. Approximate significance test for intercept smoothed over latitude and longitude: $\chi^2_{28.11} = 452.7052$ ($p < .001$). N_{total}=36,747 (N_{air}=16,211 air polluters and $N_{hazardous}$=20,536 hazardous waste large quantity generators). AIC=42358.

8 Spatial Correlation Diagnostics

Could states' tendency to free ride be contagious, such that a state is more likely to push air polluters downwind if it has neighboring states that use the same strategy? Could environmental policy be contagious, with neighboring states scoring similarly on the green index of environmental policy that we consider? In Table A.7, we consider whether there could be spatial correlation among neighboring states both in free riding and in environmental policy. To assess this, we compute two common measures of spatial association—Moran's $\mathcal{I}$ and Geary's $\mathcal{C}$—each of which asks if states' values of the variable tend to be similar to neighboring states' values. (Neighbors are defined by a shared border.) The table reports the value of $\mathcal{I}$ and of $\mathcal{C}$ for each of the three variables, showing a p-value for the statistic underneath. As can be seen at a glance, there is no spatial correlation among neighboring states in these variables.

Table A.7: Diagnostics of Spatial Correlation in Three Variables

Covariate	Moran's $\mathcal{I}$	Geary's $\mathcal{C}$
Logged ratio of air distance	−0.2842	1.2614
to hazardous waste distance	(0.9969)	(0.9943)
State average of model residual	−0.1022	1.0621
	(0.7981)	(0.7289)
Green index	0.0536	0.9283
	(0.2229)	(0.2360)

Notes: N=48. Cell entries are test statistics, p-values in parentheses. Estimates computed with R 3.2.0.

Turning to specific variables in Table A.7, we first show the spatial correlation of a logged ratio of a state's average distance of an air polluter to the downwind border over the state's average distance of a large quantity generator of hazardous waste to the downwind

border. This measure is meant to capture the unmodeled average level of free riding by a state. As can be see, Moran's $\mathcal{I}$ is negative and Geary's $\mathcal{C}$ is greater than 1: Both of these scores would actually imply a small amount of *negative* spatial correlation (e.g., an oscillating pattern wherein free riding states are likely to neighbor non-free riding states). Neither statistic is significant, so we see no evidence of state-wise spatial correlation. The second row looks at the average deviance residual from our primary model from Table 1 in the main text in order to see if correlation in states' behavior emerges once the model is considered. It is important to bear in mind that the smoothed baseline intercept in itself captures a spatial process because it forces nearby values to take on similar relative risks. Since the smoothed intercept is significant, there is spatial correlation in the process through that. However, there is not additional correlation among neighboring states beyond that, as $\mathcal{I}$ and $\mathcal{C}$ are still indiscernible. Lastly, we consider whether there may be spatial correlation in environmental policy. In this case, $\mathcal{I}$ is positive and $\mathcal{C}$ is less than 1, so both of these scores would imply a small amount of positive spatial correlation. However, the statistics are still indiscernible. So although geographic patterns are central to our project—seeing that downwind distance does predict the relative risk of an air polluter and nearby locales have a similar relative risk of air polluters—we do not see a geographic pattern of correlated state-level effects layered onto this.

9 Data Sources

Our data were drawn from the following sources:

- Geospatial data on air polluter sites: The EPA's *Geospatial Data Access Project*. Includes the name and address of each polluter, and coordinates for latitude and longitude for the major air polluter. Accessed April 17, 2012 from the Geospatial Data Download Service (http://www.epa.gov/enviro/geo_data.html).

- Geospatial data on hazardous waste generator sites: EPA *Geospatial Data Access Project*. Includes the EPA's geospatial data on large quantity generators (LQG) of hazardous waste (which we use as the control group in our article) as well as hazardous waste treatment, storage, and disposal facilities (TSDF). Data include whether the facility is a LQG or TSDF, the name and address of each facility, and coordinates for latitude and longitude for the facility. Accessed July 5, 2012 from the Geospatial Data Download Service (http://www.epa.gov/enviro/geo_data.html).

- Toxic emissions by air polluters: The EPA's 2010 Toxics Release Inventory (TRI). Accessed February 7, 2013 from the TRI program (http://www2.epa.gov/toxics-release-inventory-tri-program).

- State environmental policy stringency: Hall and Kerr's (1991) Green Index, which rates states on 67 state policy initiatives in a variety of environmental arenas, including air, water, and hazardous waste. The index includes indicators such as the sanctions available to the appropriate agencies in each state, the size of the state's pollution monitoring program, the size of the state's program budget, and a variety of specific policy indicators. The Green Index contains some indicators of state administrative assumption of federal environmental policies, which have been shown to follow a different pattern than state-initiated environmental programs. Therefore the index has been recalculated to exclude the indicators of federal program assumption.

- Locational economic development: Saiz and Clarke's (2013) measure of a state's propensity to engage in locational strategies for pursuing economic development, which is based on a content analysis of over a thousand state-level programs. These traditional "smokestack chasing" policies seek to reduce the costs of doing business relative to other states, in contrast to infrastructure and entrepreneurial approaches to economic development (Saiz & Clarke 2013). Our analyses employ an average of their index from 1986-2006.

- Environmental interest density: Operationalized as the average number of environmental interest groups in the state in 1990 and 1997. Data for the number of organized groups were compiled from lobbying registration rolls by Gray and Lowery (2001).

- Power plant site location and year of going online: The EPA's *Emissions & Generation Resource Integrated Database*. Accessed November 17, 2013 from eGRID (http://www.epa.gov/cleanenergy/energy-resources/egrid/).

- Wind direction: NOAA document entitled, "Climatic Wind Data for the United States." This document contains the prevailing wind direction data from 1930-1996 that we use in the wind kriging model. The document is dated November 1998 and was provided in personal communication with Jeff Robel on March 8, 2013.

- Maps of state borders: U.S. Census TIGER/Line Shapefiles of the American states. Accessed June 19, 2012 from the Census website (http://www2.census.gov/geo/tiger/TIGER2011/STATE/).

References

Gray, Virginia & David Lowery. 2001. "The Expression of Density Dependence in State Communities of Organized Interests." *American Politics Research* 29(4):374–391.

Hall, Bob & Mary Lee Kerr. 1991. *1991-1992 Green Index: A State-by-State Guide to the Nation's Environmental Health*. Washington, DC: Island Press.

Hedge, David M. & Michael J. Scicchitano. 1994. "Regulating in Space and Time: The Case of Regulatory Federalism." *Journal of Politics* 56:134–53.

Lowry, William R. 1992. *The Dimensions of Federalism: State Governments and Pollution Control Policies*. Durham, NC: Duke University Press.

Neyman, J. 1937. "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability." *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 236(767):333–380.

Saiz, Martin & Susan E. Clarke. 2013. Economic Development and Infrastructure Policy. In *Politics in the American States: A Comparative Analysis*, ed. Virginia Gray, Russell L. Hanson & Thad Kousser. 10th ed. Thousand Oaks, CA: CQ Press.

Wood, B. Dan. 1992. "Modeling Federal Implementation as a System: The Clean Air Case." *American Journal of Political Science* 36:40–67.