Targeting distributional impacts in the presence of behavioral responses: Lessons from maritime emissions regulation

Jamie Hansen-Lewis Michelle Marcus*

May 12, 2025

Abstract

Targeting distributional impacts is gaining importance in the design of environmental policy. To achieve this, policy makers are adopting advances in air transport models to predict the benefits of air emissions regulation. These models offer policy makers accuracy in the spatial distribution of ambient air quality improvements for a given emissions reduction, but do not take into account behavioral responses to environmental policies. We consider how the failure to account for behavioral responses when making policy predictions may have important implications for the ultimate distributional impact of such policies. We compare the distributional impacts of maritime emission regulation predicted from the policy maker's air transport model to the realized distributional impacts. We then decompose the prediction error from two components: model error, whereby the predictions of air transport models fail to account for behavioral responses of polluting firms, and sorting error, whereby the targeted population migrates.

*Hansen-Lewis is Assistant Professor in the Department of Agricultural and Resource Economics at the University of California, Davis (email: jhansenlewis@ucdavis.edu). Marcus is Associate Professor in the Department of Economics at Vanderbilt University and a Research Associate at the National Bureau of Economic Research (email: michelle.marcus@vanderbilt.edu). We gratefully acknowledge financial support for this project from the Transportation Economics in the 21st Century Initiative of the National Bureau of Economic Research and the US Department of Transportation as well as the USDA National Institute of Food and Agriculture, Hatch project 1024474, and the Giannini Foundation of Agricultural Economics. All remaining errors are our own.

1 Introduction

Research has consistently shown that exposure to air pollution disproportionately affects marginalized communities (Tessum et al., 2021; Colmer et al., 2020), and addressing these inequalities has become an increasingly important goal of environmental policy (Polonik et al., 2023; Picciano et al., 2023; Goforth and Nock, 2022; Zhu et al., 2022; Currie et al., 2023). To achieve this, air transport models have become essential tools in shaping environmental policies, particularly for regulating emissions. These models offer policymakers greater accuracy in the spatial distribution of ambient air quality improvements for a given emissions reduction by incorporating factors such as weather patterns, geographic features, and chemical interactions to predict pollution dispersion. By offering more precise predictions, these models allow policymakers to better understand the distributional impacts of policies, including how pollution might affect marginalized communities.

Despite their strengths, air transport models may fail to perfectly predict the effectiveness or distributional impact of environmental policies. Even if they accurately capture pollution dispersion, they typically do not incorporate behavioral responses from firms and individuals in response to the policy. For instance, firms might strategically respond to regulation by relocating polluting activities to avoid regulatory costs (Tanaka et al., 2022; Becker and Henderson, 2000). These models may also overlook broader policy interactions, such as regulatory rebound, where the benefits of one policy are offset by increased emissions in another sector (Fowlie et al., 2024). Individuals may also respond by changing their behaviors, such as avoidance strategies and defensive investments, which may alter the realized impacts on health (Barwick et al., 2024; Zhang and Mu, 2018; Deschenes et al., 2017; Zivin et al., 2011). Moreover, failing to account for individual sorting patterns (Banzhaf et al., 2019) may alter the ultimate distributional impacts of environmental policies.

This paper considers how failure to account for behavioral responses may have important implications for the distributional impact of environmental policies. We study the first major policy regulating maritime pollution in the US, which required all commercial ships to operate with low-sulfur fuel within 200 nautical miles off the coast or to install abatement equipment. Hansen-Lewis and Marcus (2025) show that this policy reduced fine particulate matter and improved infant birth outcomes in coastal areas with heavy ship traffic. Despite these improvements, the ex-post impact of the policy on air quality was only about 55 percent of the ex-ante predicted declines based on the EPA's air transport model. Hansen-Lewis and Marcus (2025) find evidence consistent with firm and individual behavioral responses that altered the policy's effectiveness but were not taken into account in ex-ante predictions. In this paper, we compare the predicted distributional impacts, based on policy-makers' air transport models, with the realized distributional impacts. We then disentangle the overall prediction error into two key components: model error, such as failing to incorporate polluting firms' behavioral responses, and error driven by individual sorting patterns.

We find that the realized distributional impacts of the policy significantly diverged from initial expectations and that this divergence was primarily explained by model error rather than sorting error. Notably, Hispanic populations, who were initially projected to benefit most from the policy, experienced the largest shortfall in realized air quality improvements due to their disproportionate presence in areas where shipping firms engaged in avoidance behavior. We also find that while race/ethnicity and income influenced disparities in the gap between predicted and realized benefits, differences across age groups were minimal.

This study sheds light on how the failure to capture behavioral dynamics can distort predictions, particularly in the context of environmental justice and policy effectiveness. Ex-post evaluation of the effectiveness and distributional effects of environmental policies can help inform future policy design and predictive models to better anticipate strategic responses to regulation and to consider mechanisms to mitigate compliance avoidance. These insights contribute to the growing literature on the intersection of regulatory policy, behavioral economics, and environmental justice, offering guidance for the development of more effective and equitable pollution control strategies.

2 Policy Background

2.1 North American Emission Control Area

The North American Emission Control Area (ECA) was established in August 2012 and required the use of low sulfur fuel (up to 1.0 percent sulfur oxides) in coastal waters.[1] The

[1] The ECA regulation also tightened standards for engine emissions of nitrogen oxides, but these additional standards applied to only a small subset of ship traffic: new US-flagged ships delivered after the policy came into effect.

new standards applied to all commercial ships and distributors of marine fuel. To reduce emissions, ships could use compliant fuel or an approved equivalent method, such as a scrubber. Most ships complied with the policy by switching to low sulfur fuel, stored in one of multiple fuel tanks, when entering the regulated area.[2] Low-sulfur fuel can cost 30-50 percent more than bunker fuel, and fuel accounts for up to 75 percent of an ocean carrier's operation costs.

The regulation applied within the exclusive economic zone of the United States and Canada, as seen in Figure 1. This jurisdiction extended for 200 nautical miles from the coast except in parts of southern California, Texas, and Florida where it was equidistant from non-participating neighboring countries, Mexico, the Bahamas, and Cuba.

The U.S. Coast Guard (USCG) was responsible for enforcing compliance with ECA regulations through both scheduled and unscheduled inspections.[3] These inspections could occur during routine port state control exams, domestic vessel inspections, and vessel safety examinations. Vessel operators were required to demonstrate compliance to USCG personnel, including port state control examiners, marine inspectors, and boarding officers, who conducted inspections both in port and at sea. Operators needed to provide documentation such as fuel purchase and delivery records, fuel samples, written fuel oil changeover procedures, and a fuel oil changeover logbook. The logbook had to detail the volume of compliant fuel in each tank and record the date, time, and ship's position during any fuel oil changeover operations. Additionally, the sale of non-compliant fuel was prohibited. However, incoming ships could request an exemption from penalties if they could demonstrate that compliant fuel was unavailable.

2.2 Anticipated Distributional Impacts of the ECA

The EPA developed the Community Multiscale Air Quality Modeling (CMAQ) System as a component of their proposal to justify the ECA policy (U.S. EPA, 2009b). This model focused on predicting the decline in fine particulate matter (PM2.5) from ships as a consequence of the ECA policy.[4]

[2]Scrubber installation was uncommon except among cruise and passenger ships (Hellenic Shipping News, 2014).

[3]Violations were subject to the Act to Prevent Pollution from Ships, with penalties of up to $25,000 per violation. Each day a violation continued was treated as a separate offense.

[4]The regulation defined fuel content limits for sulfur exhaust as a means to abate fine particulate matter and protect health from both primary and secondary pollutants that are formed in the atmosphere through chemical interaction (U.S. EPA, 2017, 2009b,a, 2016).

The CMAQ model used by the EPA incorporates an advanced aerosol transport model to predict changes in air quality, offering several key benefits for predicting the ECA's effects. By isolating emissions from ships specifically, the CMAQ model allows for a more precise understanding of the contribution of maritime emissions to ambient air pollution levels. By integrating data on ship traffic patterns rather than relying solely on port locations, the model provides a more comprehensive representation of shipping emissions, including emissions from vessels operating along major shipping routes. The model's ability to simulate how emissions interact with meteorological conditions and other pollutants further enhances its predictive accuracy, and enables predictions of pollution dispersion and deposition on a regional and local scale. Figure 2 shows the predicted change in fine particulate matter from the ECA for census tracts near heavy ship traffic based on the EPA's CMAQ model. Areas with the largest predicted reductions in PM2.5 are not only near major port cities, but also along heavily trafficked shipping routes.

Based on the CMAQ predictions, Hansen-Lewis and Marcus (2025) highlight important differences in the anticipated distributional impact across race/ethnicity groups for ship emissions relative to stationary pollution sources, such as port emissions. Port emissions, for example, disproportionately affect non-white populations, consistent with a large environmental justice literature studying the distributional impacts of stationary land-based pollution sources (Cain et al., 2024; Tessum et al., 2021; Banzhaf et al., 2019). However, the anticipated distributional impact across race/ethnicity groups is quite different when focused on reducing all ship emissions, as predicted by the CMAQ model. Panel (a) of Figure 6 shows the cumulative distribution function of exposure to ship emissions across race/ethnicity groups, where the x-axis represents the anticipated decline in PM2.5 from reduced ship emissions due to the ECA. Hispanics (dotted line) are the most disproportionately exposed to ship emissions and, therefore, they are expected to benefit most from the ECA policy. Interestingly, black (dashed line) and white (solid line) individuals are similarly exposed to all ship emissions, despite the fact that black individuals live closer to ports on average. Figure 3 shows the spatial distribution of each race/ethnicity group. Hispanic populations are especially concentrated in areas expected to receive the greatest PM2.5 improvement from the policy as shown in Figure 2.

We build upon this analysis to consider the anticipated distributional impact of the ECA on other important subgroups by age and income in panels (b) and (c) of Figure 6.

In panel (b), variation in the anticipated decline in PM2.5 across age groups is much smaller than across race/ethnicity groups. While exposure to ship emissions is very similar for the population under age 5 (solid line) and between ages 20 to 65 (dashed line), there is slightly less exposure among the population over age 65 (dotted line). Figure 4 shows the spatial distribution of each age group. While the spatial distribution of the population under age 5 and age 25 to 65 is fairly uniform, the concentration of the population above age 65 is higher in Florida. Panel (c) shows that there is slightly larger exposure to ship traffic among the population with income below the federal poverty line (solid line) relative to households with income above $150,000 (dashed line). Figure 5 shows there is a higher concentration of high income households in the northeast and west coast, whereas the population below poverty is relatively more concentrated in the south. Based on these patterns, we would expect the ECA to have the greatest benefit for populations that are Hispanic, under age 65, and with income below the poverty level.

2.3 Effectiveness of the ECA

The ECA was effective in reducing air pollution and protection the health of coastal populations. Hansen-Lewis and Marcus (2025) find that the introduction of maritime emissions control areas around the US coastline led to a 4 percent decrease in the population-weighted average fine particulate matter across counties within 200km of heavy ship traffic. Consistent with the air quality improvements, they also find that the ECA led to a 1.7 percent average reduction in the incidence of low birth weight and a 2.8 percent decline in infant mortality. This translates to approximately 1,536 fewer low birth weight infants per year and 228 fewer deaths per year under age one in areas near ship traffic.[5] Using the EPA's value of a statistical life, this reduction in infant deaths translates into about $2.16 billion per year.

Hansen-Lewis and Marcus (2025) also show that using the EPA's atmospheric aerosol transport model for ship pollution, the CMAQ model, instead of distance as a proxy for exposure provided a meaningful improvement in the precision of estimation. However, they show that only about half of the anticipated decline in PM2.5 predicted by the EPA's transport model was realized after the ECA's implementation in 2012. Ultimately, Hansen-

[5]Based on the 95 percent confidence intervals of the estimates, these benefits range from 746 to 2,327 fewer low birth weight infants per year and 64 to 393 fewer deaths per year under age one.

Lewis and Marcus (2025) find evidence consistent with behavioral responses among ship operators, other polluters, and individuals that diminished the policy's impact, but were not incorporated into the CMAQ model predictions.

For example, because low-sulfur fuel was more expensive than typical bunker fuel, the implementation of the ECA created an incentive for ships to reroute to avoid the regulation in areas with reduced jurisdiction. Areas in southern California, Texas, and Florida had reduced jurisdiction, as seen in Figure 1, because Mexico, Cuba, and the Bahamas did not participate in the ECA. In fact, Hansen-Lewis and Marcus (2025) find evidence consistent with ships altering their routes to avoid using the costly low-sulfur fuel required inside the ECA. Reductions in fine particulate matter from the ECA were smaller than ex-ante predictions in areas with reduced jurisdiction, where it was easiest for ships to reroute around the ECA boundary.[6]

Given that the ex-ante predictions of the policy's impacts were not fully realized, it is important to assess whether the ex-post distributional impacts of the policy meet ex-ante expectations. Because many of the unanticipated behavioral responses by ship operators and other polluters were heterogeneous across space, it is likely that the realized distributional impacts of the ECA policy differ from the expected patterns in Figure 6.

3 Data

To conduct our analysis, we use the EPA's Community Multiscale Air Quality Modeling (CMAQ) System. The EPA developed these predictions as a component of their proposal to justify the ECA policy (U.S. EPA, 2009b). We obtained the output of the CMAQ ECA analysis in 12km resolution raster grids for (i) 2020 annual mean PM2.5 concentration under business as usual and (ii) 2020 annual mean PM2.5 concentration under the ECA regulation. The difference between (i) and (ii) represents the ex-ante expected reductions in PM2.5 from the policy. The CMAQ predictions are based on 2002 ship traffic and fleet characteristics. Traffic is scaled to approximate 2020 ship traffic levels but is not adjusted for behavioral adaptations in shipping activity as a result of the ECA regulation.

Our main air quality outcome is fine particulate matter (PM2.5). PM2.5 is both a

[6] Klotz and Berazneva (2022) and Moore et al. (2018) find similar patterns of behavioral responses for ships, showing evidence of ships altering their routes to travel just outside the California ECA to minimize use of expensive low-sulfur fuel.

direct and secondary pollutant of ship exhaust, and over-land secondary PM2.5 was the criterion pollutant targeted by the ECA fuel content regulation (U.S. EPA, 2009b). Our air quality data comes from the United States Environmental Protection Agency Air Quality System (AQS) database (U.S. EPA, 2023).

To measure the ex-post reductions in PM2.5 from the policy for each tract, we first remove regional trends and seasonality from our PM2.5 measures by controlling for region-by-year and county-by-season fixed effects. These controls will capture seasonal patterns that differ across counties and broad trends that affect all counties within a region, which are unlikely to be related to pollution reductions due to ship emissions. We average residuals from 2009 to 2011 and 2013 to 2016 to provide measures of PM2.5 before and after the policy, respectively.[7] The difference between these two measures of PM2.5 represents the ex-post realized reductions in PM2.5 from the policy.

We combine these data with census tract-level data on population characteristics. We use 5-year estimates from the American Community Survey (ACS). Our measures of demographic characteristics before and after the policy cover the 5-year periods 2007 to 2011 and 2013 to 2017, respectively. We use the percent of the population that is non-Hispanic white, non-Hispanic black, non-Hispanic other race, Hispanic, under age 5, age 20-65, over age 65, below the federal poverty line, and with household income over $150,000.

We restrict our sample to a balanced panel of air quality monitors that are observed at least once per year from 2009 to 2016. In our main results, we assign each tract in our sample to the closest air quality monitor within 50km. About 93 percent of tracts have at least one monitor within 50km. We focus on the sample of census tracts within counties where the population-weighted centroid is within 200km of heavy ship traffic. Table 1 shows the characteristics of our analysis sample in column (2) relative to the full sample in column (1). Across race/ethnicity, age, and income groups, our analysis sample is very similar to the full sample.

[7] We exclude observations in 2012 as it is a partially treated year.

4 Methods

There are two main pathways that could lead to differences in the intended and realized distributional impacts of the policy. The first pathway is that the policymaker might incorrectly estimate the spatial distribution of air quality improvements from the policy. The second pathway is that the spatial distribution of the population of interest could change due to sorting. Our objective is to first characterize the overall difference between the intended and realized distributional impacts of the ECA policy and then separate the relative contributions of each pathway.

Our first set of results compares the intended and realized distributional impacts of the ECA policy. For each demographic group, we estimate the fraction of the sample in the 2011 ACS data that obtained at least the mean improvement from the policy, roughly 0.4 μgm^{-3}, using the CMAQ predictions for each tract. This represents the anticipated distributional impact of the policy prior to implementation. We then compare this estimate with the fraction of the sample in the 2017 ACS data that obtained at least a 0.4 μgm^{-3} realized improvement in PM2.5. The difference in these fractions indicates how much the CMAQ model misrepresented the actual improvement in PM2.5 for each demographic group when both error pathways are present. In addition to summarizing differences at the mean predicted improvement, we plot the cumulative distribution functions of the fraction of each demographic group obtaining each level of predicted PM2.5 improvement based on the 2011 ACS data. We compare these to the cumulative distribution functions of the fraction of each demographic group obtaining each level of actual PM2.5 improvement based on the 2017 ACS data.

Our second set of results considers the degree to which the gap between intended and realized distributional impacts arises from model error. Differences between CMAQ-predicted PM2.5 and actual PM2.5 may arise from behavioral responses, such as the atmospheric model failing to incorporate firm responses when predicting pollution declines from the policy, but also from any way in which the model imperfectly predicted PM2.5 in levels and spatial distribution, such as from an inaccurate inventory of emissions or simulation of the atmosphere. We highlight the differences between intended and realized improvements with the population held fixed in the 2011 spatial distribution. Specifically, we compare the fraction of the sample in the 2011 population distribution that obtained at least a 0.4 μgm^{-3} improvement in PM2.5 to the fraction of the sample in the 2011 population distribution that the CMAQ model predicted to obtain at least the same improvement in PM2.5. In addition, we plot the cumulative distribution functions of the fraction of each demographic group obtaining each level of predicted and realized PM2.5 improvement, holding the spatial distribution of the population constant in 2011. Last, to highlight the differences in model error across groups, we we calculate for each tract the difference between actual and predicted PM2.5 and plot the CDFs of the difference.

Our final set of results considers how changes in the spatial distribution of the populations of interest contributed to the overall gap between intended and realized distributional impacts. We highlight differences between intended and realized improvements assuming CMAQ PM2.5 improvements were correct. Specifically, for each demographic group, we compare the fraction of the sample in the 2017 ACS data that CMAQ predicted to obtain at least the mean realized improvement in PM2.5 to the fraction of the sample in the 2011 ACS data that CMAQ predicted to obtain at least the mean realized improvement in PM2.5. To further illustrate the role of population sorting, we plot for each demographic group the cumulative distribution functions of the fraction of each demographic group obtaining each level of predicted PM2.5 improvement based on the 2011 and 2017 population distributions.

5 Results

5.1 Overall error

First, we compare the intended and realized distributional impacts of the ECA policy. This comparison summarizes the overall error, which can arise from both model misspecification ("model error") and population sorting ("sorting error"). Column (1) of Table 2 shows the percent of the population in each demographic group based on the 2011 ACS that was expected to obtain at least the mean PM2.5 improvement from the policy, roughly 0.4 μgm^{-3}, using the CMAQ predictions for each tract. For example, about 50.3 percent of the black population was expected to receive a reduction in PM2.5 of at least 0.4 μgm^{-3} from the policy. The policy was expected to improve air quality the most for Hispanic and other race populations. About 69.5 percent of the Hispanic population and 63.2 percent of the other race population were expected to receive a PM2.5 improvement of 0.4 μgm^{-3} or greater. Expected improvements for the white population was the lowest with only

43.2 percent expected to receive at least the mean improvement in PM2.5 from the policy. Column (2) reports the percent of the population in each group that actually obtained at least the mean PM2.5 improvement from the policy, based 2017 ACS data and the realized PM2.5 improvement from the policy. Column (3) reports the overall error, or the difference between the expected and actual PM2.5 improvements shown in columns (1) and (2).

Consistent with Hansen-Lewis and Marcus (2025), we see that the realized PM2.5 improvements in column (2) were below the expected improvements in column (1) for all demographic groups. Interestingly, the difference between intended and realized PM2.5 declines was not uniform across different demographic groups, as seen in column (3). The biggest differences arise across race/ethnicity groups. For example, the gap between intended and realized PM2.5 declines was the smallest for the white population, only about 10 percentage point difference. Although only 43 percent of the population was expected to obtain at least the mean PM2.5 improvement, about 33 percent of the population did obtain at least the mean PM2.5 improvement for the white population. The gap between intended and realized PM2.5 declines was about 13 percentage points for the black population and 18 percentage points for the other race population. The Hispanic population had the largest gap–the realized PM2.5 decline was about 30 percentage points lower than the expected PM2.5 decline. While 69.5 percent of the Hispanic population was expected to receive at least the mean PM2.5 improvement, only 40 percent actually did.

Differences across age groups are more modest. The predicted improvement in PM2.5 from the policy in column (1) is fairly similar for populations under age 5, age 20-65, and over age 65. The predictions range from 50.2 percent to 53.5 percent of the population expected to receive at least the mean improvement in PM2.5 from the policy, with the highest value for the youngest group, under age 5. Because realized improvements in PM2.5, in column (2), are also similar across age groups, there is little variation in the overall error by age shown in column (3). The overall error ranges from 14.8 to 16.3 percentage points and is the largest for the very young.

Last, we also observe differences across income groups. In column (1), we see that low income households were expected to benefit most from the policy, with 55.9 percent of the population below poverty expected to receive at least the mean improvement in PM2.5 from the ECA. Only 49.7 percent of the high income population, with income over \$150,000, was expected to receive the same improvement. However, the realized improvements in PM2.5 from the policy in column (2) were very similar across low and high income groups, 36.7 and 37.8 percent, respectively. Overall error in column (3) for the below poverty population was therefore around 19.2 percentage points, which was higher than the overall error for the high income population, which was about 11.9 percentage points.

While Table 2 summarizes the key differences across groups at a single point in the distribution, Figures 7, 8, and 9 summarize the error for the whole distribution. In each figure, panel (a) documents the overall error for each race/ethnicity, age, or income subgroup. For example, the solid line in row 2 of Figure 7a plots the fraction of the black population, based on 2011 ACS data, that was expected to obtain each level of PM2.5 improvement from the policy, or greater. For a PM2.5 change of -0.4 μgm^{-3} on the x-axis, which is the reference value for Table 2, the value of the solid line corresponds to the value in column (1) of Table 2, which is 50.3 percent.

Because the CMAQ model predicted a decline in PM2.5 from the policy for all locations, we can see that 100 percent of the black population was expected to obtain a decline in PM2.5. The solid line reaches 1 at a value slightly less than zero on the x-axis. In contrast, the realized change in PM2.5 was not always negative, meaning some locations saw an increase in PM2.5. The dotted line shows the fraction of the black population, based on 2017 ACS data, that actually received each level of PM2.5 improvement or greater. For a PM2.5 change of -0.4 μgm^{-3} on the x-axis, the value of the dotted line corresponds to the value in column (2) of Table 2, which is 37.2 percent.

The gap between the solid and dotted lines shows the overall difference between expected and actual PM2.5 improvements, similar to comparing columns (1) and (2) in Table 2. Figures A1 to A3 in the appendix show the probability density functions. For example, in row 2 of Figure A1a, the solid (dotted) line shows the fraction of the black population, based the 2011 (2017) ACS data, receiving each level of predicted (actual) PM2.5 improvement.

A few patterns are notable and consistent with the results from Table 2. First, we notice that the expected PM2.5 improvements (solid lines) are shifted to the left relative to the actual PM2.5 improvements (dotted lines). In other words, for almost all populations, the actual improvements in PM2.5 were smaller in magnitude than the CMAQ model predictions with some populations even seeing an increase in PM2.5. This is also consistent with Hansen-Lewis and Marcus (2025) who find that only about half of predicted air quality improvements were realized by the policy. This pattern is persistent across race/ethnicity groups, age, and income groups.

Second, we observe the largest gap between predicted and realized PM2.5 improvements for the Hispanic population, consistent with Table 2. In row 3 of Figure 7a, there is a large gap between the predicted (solid line) and realized (dotted line) fraction of the Hispanic population receiving each level of PM2.5 improvement. This gap is much smaller in row 1 of Figure 7a for the white population, for example.

Third, the gap in predicted and realized PM2.5 improvements is fairly similar across age groups. Comparing the overall error for the under 5, 20-65, and over 65 populations in panel (a) of Figure 8, we observe little differences. The gap between the predicted (solid line) and realized (dotted line) PM2.5 improvement is perhaps slightly larger among the under 5 population, consistent with Table 2.

Finally, we observe that gaps by income had less variance than race/ethnicity gaps, but still show a larger difference in predicted and realized PM2.5 improvements for low income relative to high income groups. Consistent with column (3) of Table 2, the gap between the predicted (solid line) and realized (dotted line) PM2.5 improvement is larger for the below poverty population in row 1 relative to the population with income over $150,000 in row 2.

5.2 Model error vs. Sorting error

Next, we consider the degree to which the gap between intended and realized PM2.5 declines arises from error in the CMAQ model or error driven by residential sorting. First, we isolate model error by holding constant the spatial distribution of the population in 2011 and comparing the distributional impacts of the policy under the CMAQ predictions and realized air quality improvements. In Table 2, Column (4) reports the percent of the population in each demographic group based on the 2011 ACS that obtained at least the mean realized PM2.5 improvement from the policy. Comparing columns (1) and (4) isolates the difference in intended and realized pollution declines stemming from model error in the CMAQ model. Column (5) reports the difference between columns (1) and (4), the "model error". Figures 7, 8, and 9 also document the distribution of model error in panel (b). The solid (dotted) line reports the fraction of the 2011 population receiving each level of PM2.5 improvement or more based on the CMAQ model predictions (actual PM2.5 change).

Second, we isolate sorting error by holding the CMAQ model predictions constant and comparing the distributional impacts of the policy for the 2011 and 2017 population distributions. Column (6) of Table 2 reports the percent of the population in each demographic group based on the 2017 population distribution that would have obtained at least the mean PM2.5 improvement if the CMAQ predictions had been accurate. Comparing columns (1) and (6) isolates the difference in intended and realized pollution declines coming from population sorting before and after the policy. Column (7) reports the difference between columns (1) and (6), the "sorting error". Figures 7 to 9 also document the distribution of sorting error in panel (c). The solid (dotted) line reports the fraction of the 2011 (2017) population receiving each level of anticipated PM2.5 improvement based on the CMAQ model predictions.

For all demographic groups, model error appears to be a much larger driver of overall error than sorting error. In Table 2, error in the CMAQ model predictions shown in column (5) is much larger for all demographic groups than error driven by population sorting shown in column (7). Similarly, comparing panels (b) and (c) of Figures 7 to 9 we observe a much larger gap between the solid and dashed lines when isolating model error in panel (b) as opposed to sorting error in panel (c).

Patterns for model error are similar to the patterns for overall error across race/ethnicity, age, and income groups. In Table 2, white populations had the lowest gap between intended and realized pollution based on model error, about 10 percentage points, whereas Hispanic populations had the largest gap due to model error, about 29 percentage points. Model error is less variable across age and income groups. However, we do observe relatively larger error for low income relative to high income populations.

These patterns are also observed in Figures 10, which show the cumulative distribution function of model error for each race/ethnicity group in panel (a), age group in panel (b), and income group in panel (c).[8] Each line reports the fraction of the 2011 population with at least each level of model error or less. Note that positive values of the model error on the x-axis indicate that the realized PM2.5 decline was smaller than the CMAQ

[8] Figure A4 shows the probability density function of model error for each race/ethnicity, age, and income group.

model predicted. The Hispanic population (dotted line) is shifted furthest to the right in Figure 10a, indicating that the CMAQ model over-predicted PM2.5 improvements the most for this group. Differences in model error across age groups appear relatively small in Figure 10b. In Figure 10c, the below poverty group (solid line) is shifted furthest to the right, showing that the CMAQ model over-predicted PM2.5 improvements more for low income groups relative to high income groups.

When we isolate the difference between intended and realized pollution levels driven by sorting in column (7), the gaps are much smaller, less than 1 percentage point for all demographic groups. Although small in magnitude, the patterns across demographic groups are consistent with the environmental justice literature. For all race/ethnicity groups except white, we see that sorting between 2011 and 2017 led to a decrease in the percent of the population obtaining at least the mean PM2.5 improvement. In other words, a greater percent of the non-white population would have received at least the mean PM2.5 improvement from the policy if the spatial distribution of population would have been held constant at 2011. On the other hand, white populations saw an increase in the percent obtaining at least the mean PM2.5 improvement due to sorting. These results suggest that white individuals sorted into areas that were predicted to benefit most from the policy, whereas non-white individuals sorted away from these areas. The finding that population sorting plays a minor role in the overall error is consist with patterns showing that population changes relatively slowly over time within census tracts. In the Appendix, we consider a longer time horizon with 2023 ACS data and find larger shifts in population (Table A1) which correspond to greater sorting error (Table A2).

5.3 Understanding Heterogeneity in Model Error

The previous section highlighted large variation in model error across race/ethnicity groups. We now explore what may have driven this variation.

First, it is important to consider the spatial distribution of the population by race/ethnicity. Figure 3 shows the spatial distribution of each race/ethnicity group. Relative to the other groups, Hispanics are especially concentrated in coastal areas with heavy ship traffic, such as southern California, Texas, and southern Florida. Comparing these figures to the CMAQ predicted declines in PM2.5, in Figure 2, it is not surprising that the Hispanic population was expected to receive the greatest PM2.5 improvements from the ECA policy.

However, Hansen-Lewis and Marcus (2025) find evidence consistent with ships altering their routes to avoid using the costly low-sulfur fuel required inside the ECA. Such behavioral responses were not captured by the CMAQ model and are an important contributor to the model error. They show that reductions in fine particulate matter from the ECA were smaller than ex-ante predictions in areas with reduced jurisdiction, such as southern California, Texas, and Florida, where it was easiest for ships to reroute around the ECA boundary, as seen in Figure 1. These areas with reduced jurisdiction also correspond to areas with large Hispanic populations, as seen in Figure 3c. Figure 11 further highlights this correlation by plotting the distance from county centroids to the ECA boundary on the x-axis and percent of the population in different race/ethnicity groups on each y-axis. Hispanics are the only group with a negative correlation, indicating that in areas closest to the ECA boundary–where the ECA had reduced jurisdiction under 200nm–the percent Hispanic is higher. Similarly, column (8) of Table 2 reports the percent of each demographic group living in a county near these reduced jurisdiction areas. Hispanic populations are about twice as likely as other race/ethnicity groups to live in reduced jurisdiction areas.

Because Hispanic populations are especially concentrated in areas like southern California, Texas, and Florida, where the behavioral response among ships most deviated from the CMAQ model predictions, it is perhaps not surprising that the gap between predicted and realized pollution levels due to model error was highest among the Hispanic population.

6 Implications for Policy

6.1 Targeting Disparities is Difficult

Market-based policies, such as cap-and-trade, are often lauded for their cost-effectiveness in reducing overall pollution, but they are frequently criticized for failing to adequately address environmental justice goals (Sheriff, 2024; Cushing et al., 2018; Boyce and Pastor, 2013; Fowlie et al., 2012; Muller and Mendelsohn, 2009). One of the primary concerns is that these policies can lead to the persistence or even exacerbation of pollution "hotspots"—-areas with concentrated emissions that disproportionately impact low-income or minority communities. The flexibility inherent in market-based systems, which allows firms to trade emissions allowances or purchase offsets, can inadvertently reinforce geographic disparities in air quality. For communities already burdened by pollution, the lack of direct emission reductions at the local level can undermine environmental justice objectives.

An alternative to market-based policies is the implementation of uniform technology standards, which could result in more equal reductions in pollution across different areas. Such standards could better align with environmental justice goals by ensuring that all communities benefit from improved air quality. However, our analysis highlights that even uniform policies can have unintended distributional consequences. For example, in the context of the ECA, incomplete geographic coverage of a policy created unanticipated behavioral responses, which altered the socioeconomic distribution of benefits. These findings emphasize that while uniform standards may address some equity concerns, their effectiveness depends on comprehensive policy design and enforcement.

While much interest has focused on the distributional consequences of pollution exposure along the dimensions of race/ethnicity and income, distributional consequences by age are also important given a large body of evidence documenting the largest health consequences from pollution exposure among the very young and very old. For this reason, targeting vulnerable populations by age may also desirable. However, this approach poses challenges because age demographics tend to be more uniformly distributed across geographic areas, as shown in Figure 4. Despite heterogeneity in the effectiveness of the ECA across space, we saw little variation in predicted or realized improvements across age groups. Policies aiming to target specific age groups may require additional layers of granularity, such as focusing on schools, daycare centers, and elderly care facilities, to ensure meaningful protection for these sensitive groups.

6.2 Policy Takeaways

A key lesson from the implementation of the ECA policy is the importance of ex-post evaluation to determine whether policy expectations are met. Such evaluations provide valuable insights into the actual outcomes of policy interventions, including their distributional impacts, and can help refine future policy design. For example, detailed analyses of the ECA policy's effects on pollution disparities could inform adjustments to better align with environmental justice objectives.

In the specific case of the North American Emission Control Area, the policy's distributional consequences might have more closely matched expectations if certain structural and enforcement issues had been addressed. First, broader geographic participation, including countries like Mexico, Cuba, and the Bahamas, would have reduced opportunities for behavioral responses, such as rerouting shipping traffic to avoid compliance. Second, improved enforcement, such as increased inspections of ships outside of port locations, could have ensured more consistent adherence to emissions standards.

These lessons underscore the complexity of designing environmental policies that are both effective and equitable. Policymakers must carefully balance the trade-offs between cost-efficiency and equity, while ensuring that the policies are adaptable and enforceable. By incorporating rigorous evaluation mechanisms and considering broader geographic and demographic factors, future policies can better address both environmental and social objectives.

7 Conclusion

This paper examines the limitations of air transport models in predicting the distributional impacts of environmental policies, particularly when these models fail to account for behavioral responses from both firms and individuals. Using the case of the North American Emission Control Area, we demonstrate how behavioral responses and population sorting patterns led to discrepancies between predicted and realized distributional impacts from air quality improvements. We disentangle the degree to which the relocation of pollution by firms and demographic shifts contributed to these gaps. These findings underscore the importance of ex-post policy evaluation and considering behavioral dynamics in policy design to better anticipate the real-world distributional effects of environmental policies. Addressing these gaps is crucial for designing more effective and equitable environmental policies.

References

Banzhaf, Spencer, Lala Ma, and Christopher Timmins, "Environmental justice: The economics of race, place, and pollution," *Journal of Economic Perspectives*, 2019, *33* (1), 185–208.

Barwick, Panle Jia, Shanjun Li, Liguo Lin, and Eric Yongchen Zou, "From fog to smog: The value of pollution information," *American Economic Review*, 2024, *114* (5), 1338–1381.

Becker, Randy and Vernon Henderson, "Effects of Air Quality Regulations on Polluting Industries," *Journal of Political Economy*, 2000, *108* (2), 379–421.

Boyce, James K and Manuel Pastor, "Clearing the air: incorporating air quality and environmental justice into climate policy," *Climatic change*, 2013, *120*, 801–814.

Cain, Lucas, Danae Hernandez-Cortes, Christopher Timmins, and Paige Weber, "Recent findings and methodologies in economics research in environmental justice," *Review of Environmental Economics and Policy*, 2024, *18* (1), 116–142.

Colmer, Jonathan, Ian Hardman, Jay Shimshack, and John Voorheis, "Disparities in PM2. 5 air pollution in the United States," *Science*, 2020, *369* (6503), 575–578.

Currie, Janet, John Voorheis, and Reed Walker, "What caused racial disparities in particulate exposure to fall? New evidence from the Clean Air Act and satellite-based measures of air quality," *American Economic Review*, 2023, *113* (1), 71–97.

Cushing, Lara, Dan Blaustein-Rejto, Madeline Wander, Manuel Pastor, James Sadd, Allen Zhu, and Rachel Morello-Frosch, "Carbon trading, co-pollutants, and environmental equity: Evidence from California's cap-and-trade program (2011–2015)," *PLoS medicine*, 2018, *15* (7), e1002604.

Deschenes, Olivier, Michael Greenstone, and Joseph S Shapiro, "Defensive investments and the demand for air quality: Evidence from the NOx budget program," *American Economic Review*, 2017, *107* (10), 2958–2989.

Fowlie, Meredith, Edward Rubin, and Catherine Wright, "Declining power-plant emissions, co-benefits, and regulatory rebound," 2024.

_, Stephen P Holland, and Erin T Mansur, "What do emissions markets deliver and to whom? Evidence from Southern California's NOx trading program," *American Economic Review*, 2012, *102* (2), 965–993.

Goforth, Teagan and Destenie Nock, "Air Pollution Disparities and Equality Assessments of US National Decarbonization Strategies," *Nature Communications*, 2022, *13* (1), 7488.

Hansen-Lewis, Jamie and Michelle Marcus, "Uncharted waters: Effects of maritime emission regulation," *American Economic Journal: Economic Policy*, 2025, *17* (1), 37–69.

Hellenic Shipping News, *2015 SOx Limits: Is The World Fleet Scrubbing Up?* 2014.

Klotz, Richard and Julia Berazneva, "Local Standards, Behavioral Adjustments, and Welfare: Evaluating California's Ocean-Going Vessel Fuel Rule," *Journal of the Association of Environmental and Resource Economists*, 2022, *9* (3), 383–424.

Moore, Thomas J, Jessica V Redfern, Michael Carver, Sean Hastings, Jeffrey D Adams, and Gregory K Silber, "Exploring ship traffic variability off California," *Ocean & Coastal Management*, 2018, *163*, 515–527.

Muller, Nicholas Z and Robert Mendelsohn, "Efficient pollution regulation: getting the prices right," *American Economic Review*, 2009, *99* (5), 1714–1739.

Picciano, Paul, Minghao Qiu, Sebastian D Eastham, Mei Yuan, John Reilly, and Noelle E Selin, "Air Quality Related Equity Implications of U.S. Decarbonization Policy," *Nature Communications*, 2023, *14* (1), 5543.

Polonik, Pascal, Katharine Ricke, Sean Reese, and Jennifer Burney, "Air quality equity in US climate policy," *Proceedings of the National Academy of Sciences*, 2023, *120* (26), e2217124120.

Sheriff, Glenn, "California's GHG Cap-and-Trade Program and the Equity of Air Toxic Releases," *Journal of the Association of Environmental and Resource Economists*, 2024, *11* (1), 137–170.

Tanaka, Shinsuke, Kensuke Teshima, and Eric Verhoogen, "North-South displacement effects of environmental regulation: The case of battery recycling," *American Economic Review: Insights*, 2022, *4* (3), 271–288.

Tessum, Christopher W, David A Paolella, Sarah E Chambliss, Joshua S Apte, Jason D Hill, and Julian D Marshall, "PM2.5 Polluters Disproportionately and Systemically Affect People of Color in the United States," *Science Advances*, 2021, *7* (18), eabf4491.

U.S. EPA, "Integrated Science Assessment for Particulate Matter," *US Environmental Protection Agency Washington, DC*, 2009.

_ , "Proposal to Designate an Emission Control Area for Nitrogen Oxides, Sulfur Oxides and Particulate Matter: Technical Support Document," Technical Report, Assessment and Standards Division, Office of Transportation and Air Quality, U.S. Environmental Protection Agency 2009.

_ , "Integrated Science Assessment for Oxides of Nitrogen - Health Criteria," *US Environmental Protection Agency Washington, DC*, 2016.

_ , "Integrated Science Assessment for Sulfur Oxides: Health Criteria ," Technical Report, U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment 2017.

_ , "Air Quality System (AQS) API," 2023. https://campd.epa.gov/data/custom-data-download (accessed May 8, 2023).

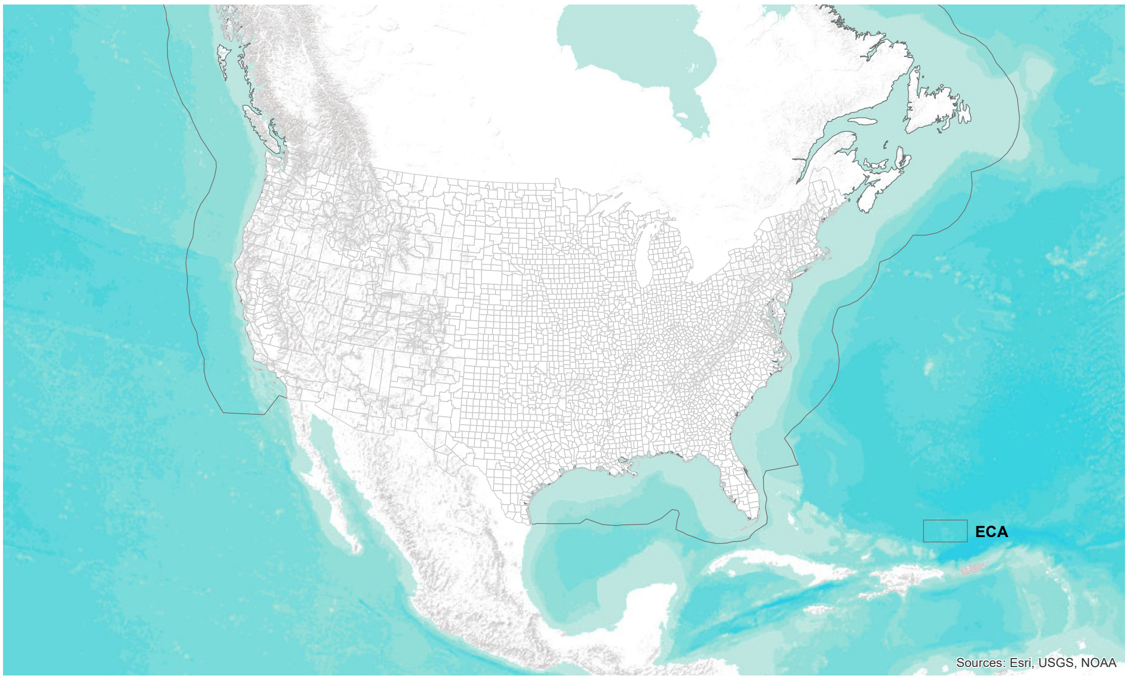
Zhang, Junjie and Quan Mu, "Air pollution and defensive expenditures: Evidence from particulate-filtering facemasks," *Journal of Environmental Economics and Management*, 2018, *92*, 517–536.

Zhu, Shupeng, Michael Mac Kinnon, Andrea Carlos-Carlos, Steven J Davis, and Scott Samuelsen, "Decarbonization Will Lead to More Equitable Air Quality in California," *Nature communications*, 2022, *13* (1), 5738.

Zivin, Joshua Graff, Matthew Neidell, and Wolfram Schlenker, "Water quality violations and avoidance behavior: Evidence from bottled water consumption," *American Economic Review*, 2011, *101* (3), 448–453.

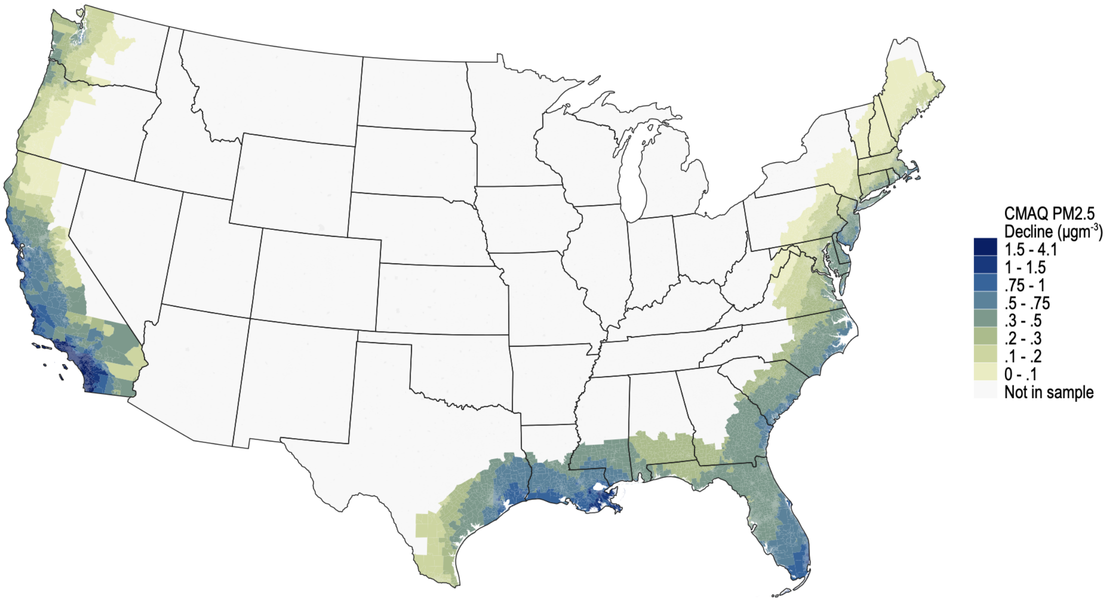
8 Figures

Figure 1: North American Emission Control Area Boundary



Note: Figure shows the regulated area for the North American Emission Control Area. Low sulfur fuel was required within the outlined boundary.

Figure 2: CMAQ Predicted Change in PM2.5 from the ECA at the Census Tract Level

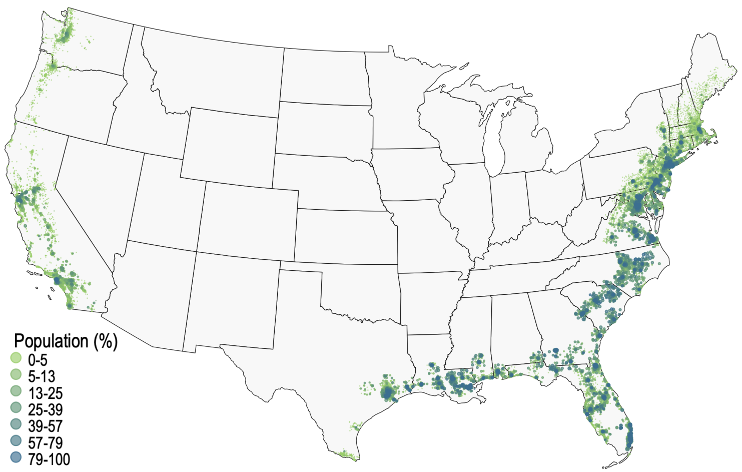


Note: Figure shows the predicted decline in fine particulate matter from implementation of the North American Emission Control Area based on the CMAQ model. Declines in PM2.5 are measured at the population-weighted centroid for all census tracts within 200km of heavy ship traffic. Darker colors indicate a greater predicted decline in PM2.5. Data are from U.S. EPA (2009b).

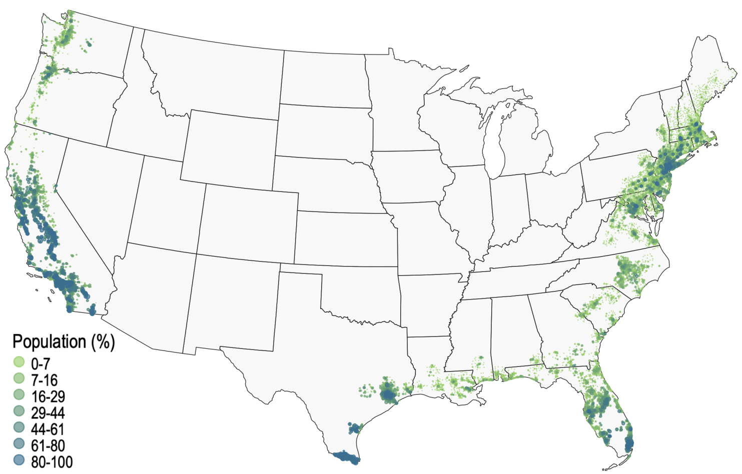
Figure 3: **Spatial Distribution of Population Across Racial Groups**

(a) *White*

(b) *Black*



(c) *Hispanic*



(d) *Other*

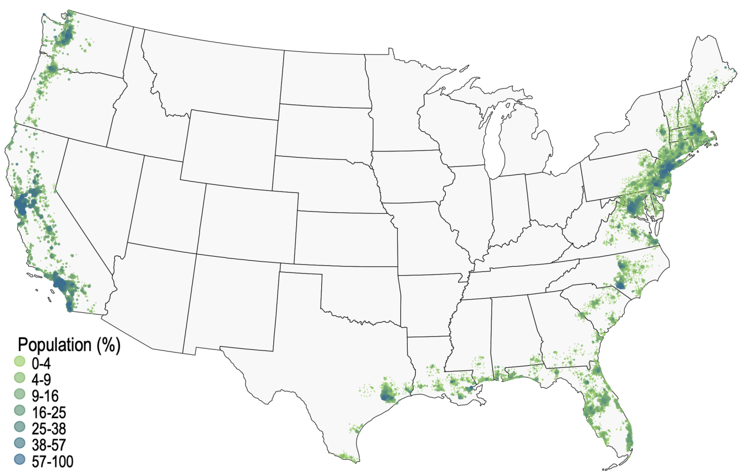
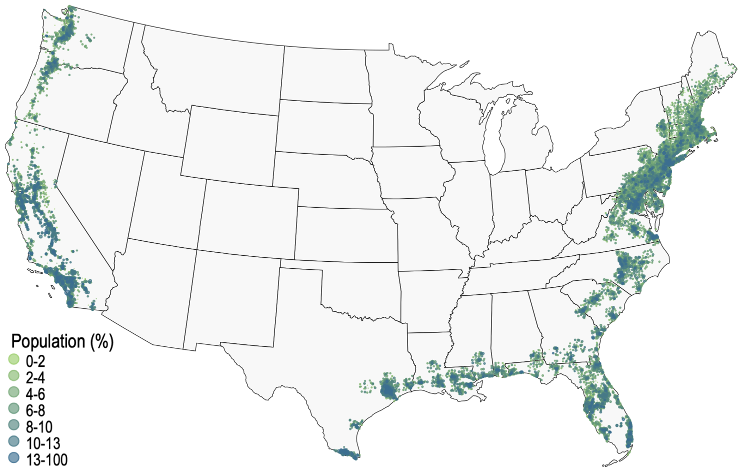
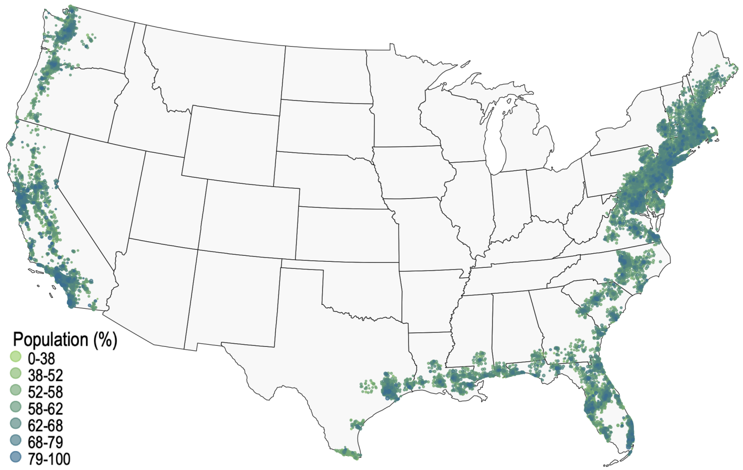


Figure 4: **Spatial Distribution of Population Across Age Groups**

(a) *Under 5*



(b) *25-65*



(c) *Over 65*

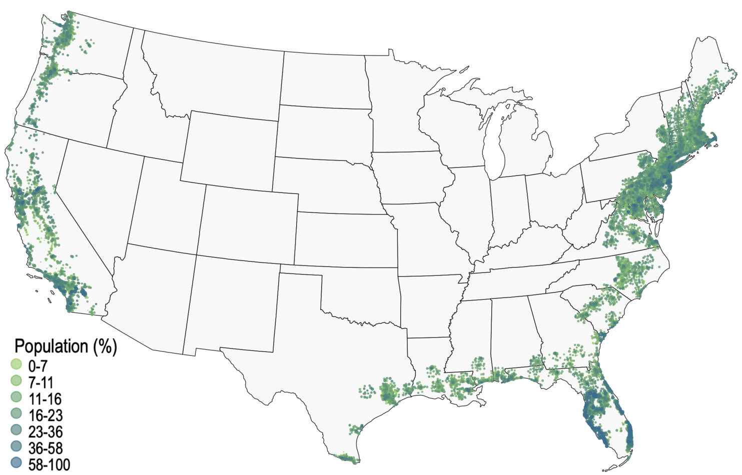
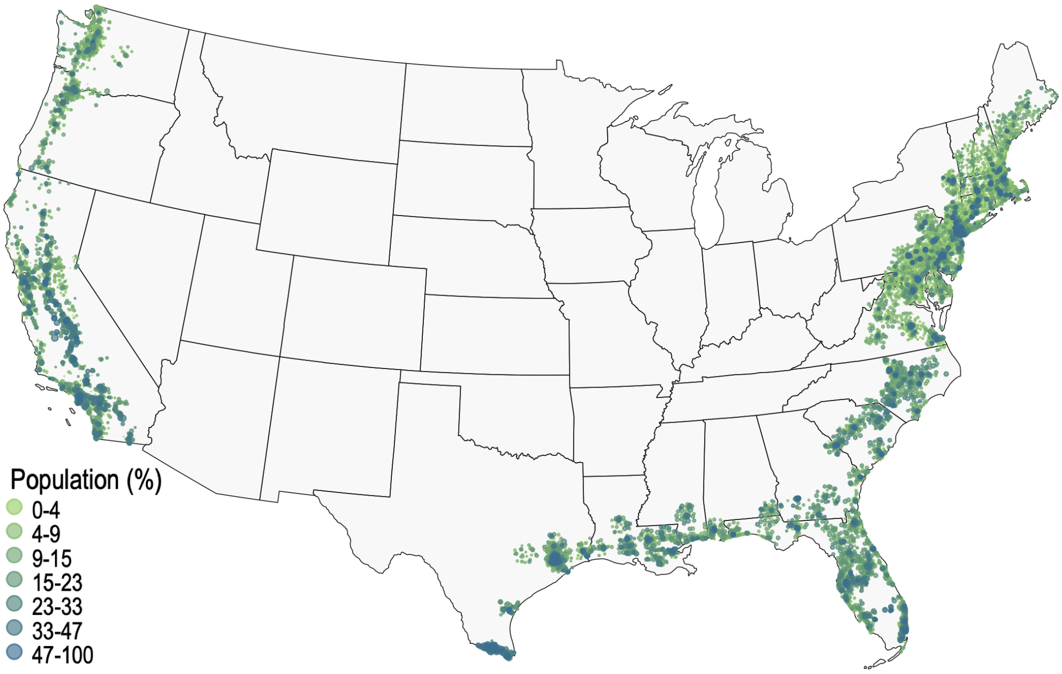


Figure 5: **Spatial Distribution of Population Across Income Groups**

(a) *Below Poverty*



(b) *Over $150k*

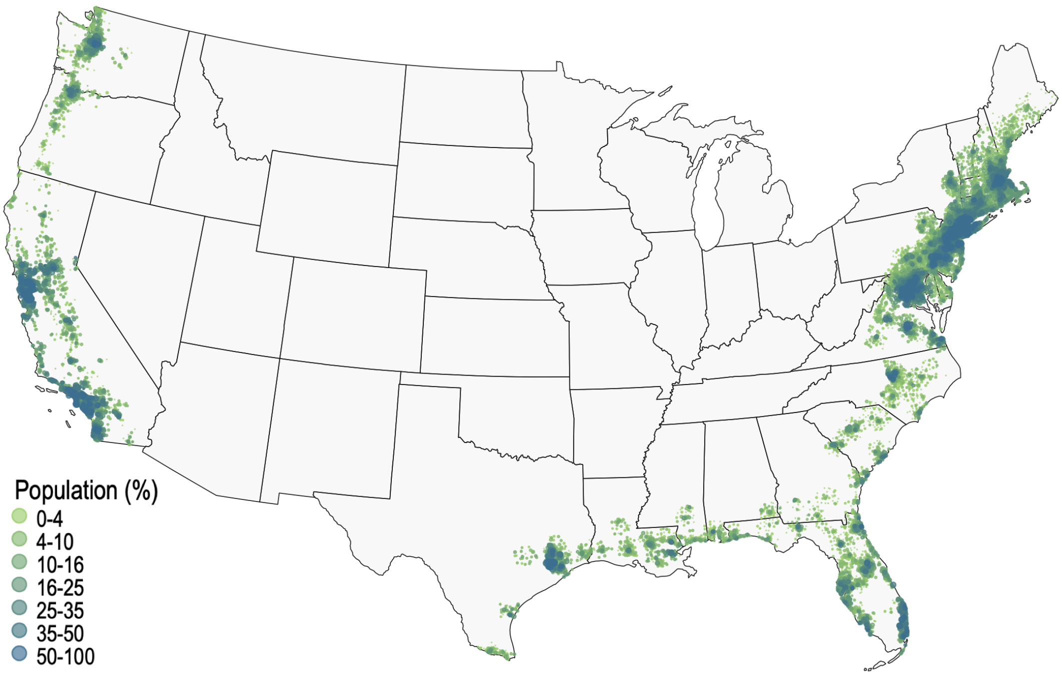
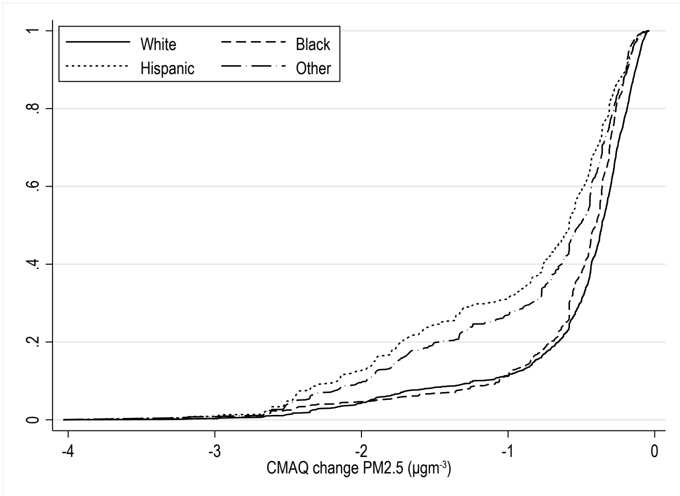
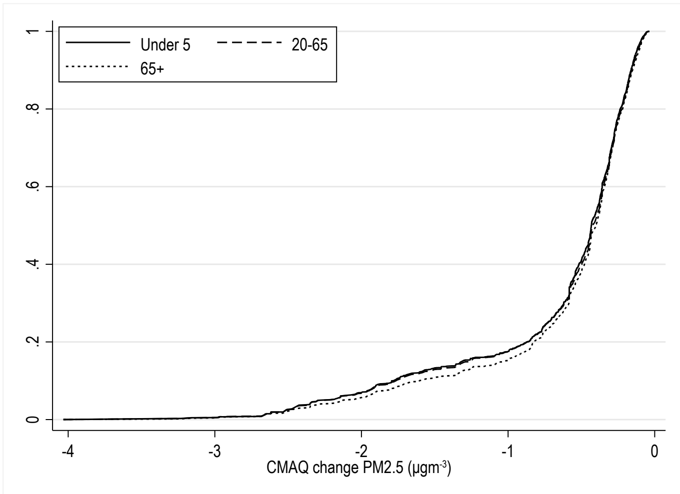


Figure 6: **Disproportionate Exposure Among Populations Exposed to Maritime Pollution**

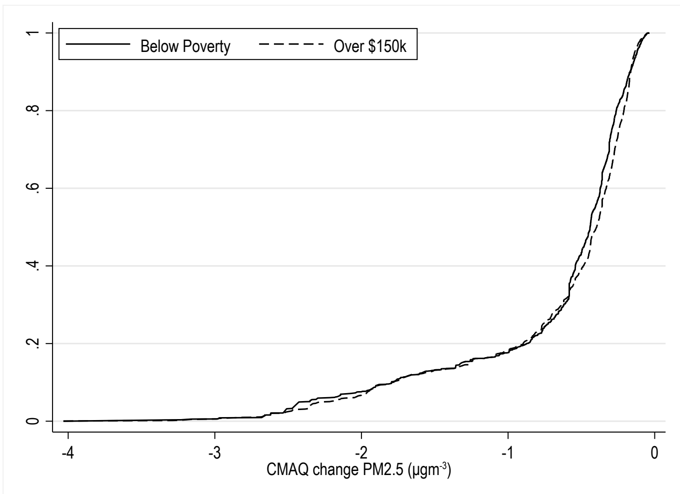
(a) *Racial Groups*



(b) *Age Groups*



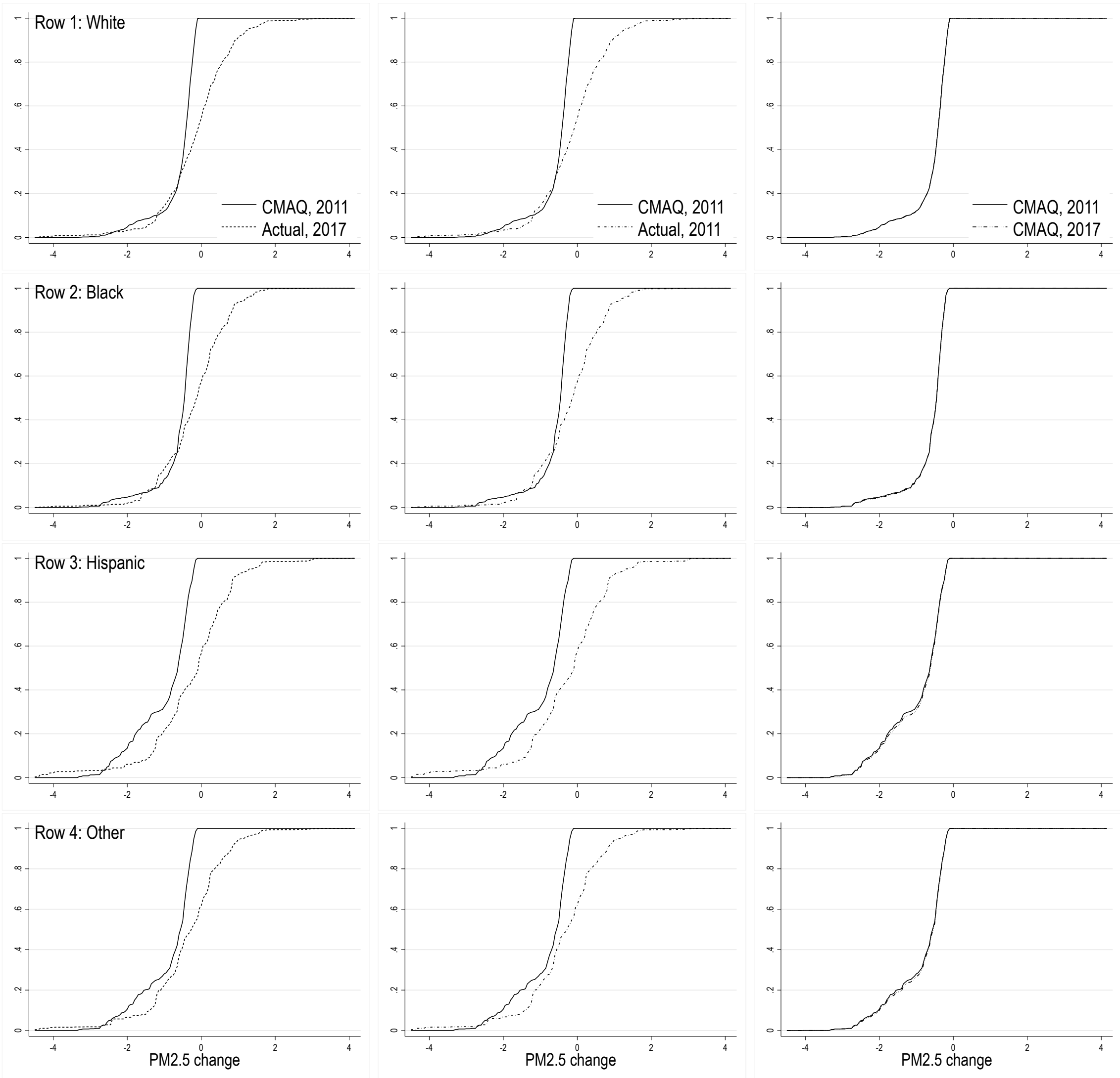
(c) *Income Groups*



Note: Demographic information from 2011 census tract data. We restrict to our analysis sample, which includes tracts within 200km of heavy ship traffic and with an air quality monitor within 50km. Panel (a) shows the cumulative distribution of individuals by race/ethnicity over intensity of ECA predicted improvement from CMAQ. Panel (b) repeats panel (a) by age group. Panel (c) repeats panel (a) by income group. The ECA predicted improvement is measured by the predicted change from requiring low sulfur maritime fuel from the CMAQ model at the centroid of each census tract.

Figure 7: **Error Comparisons by Race**

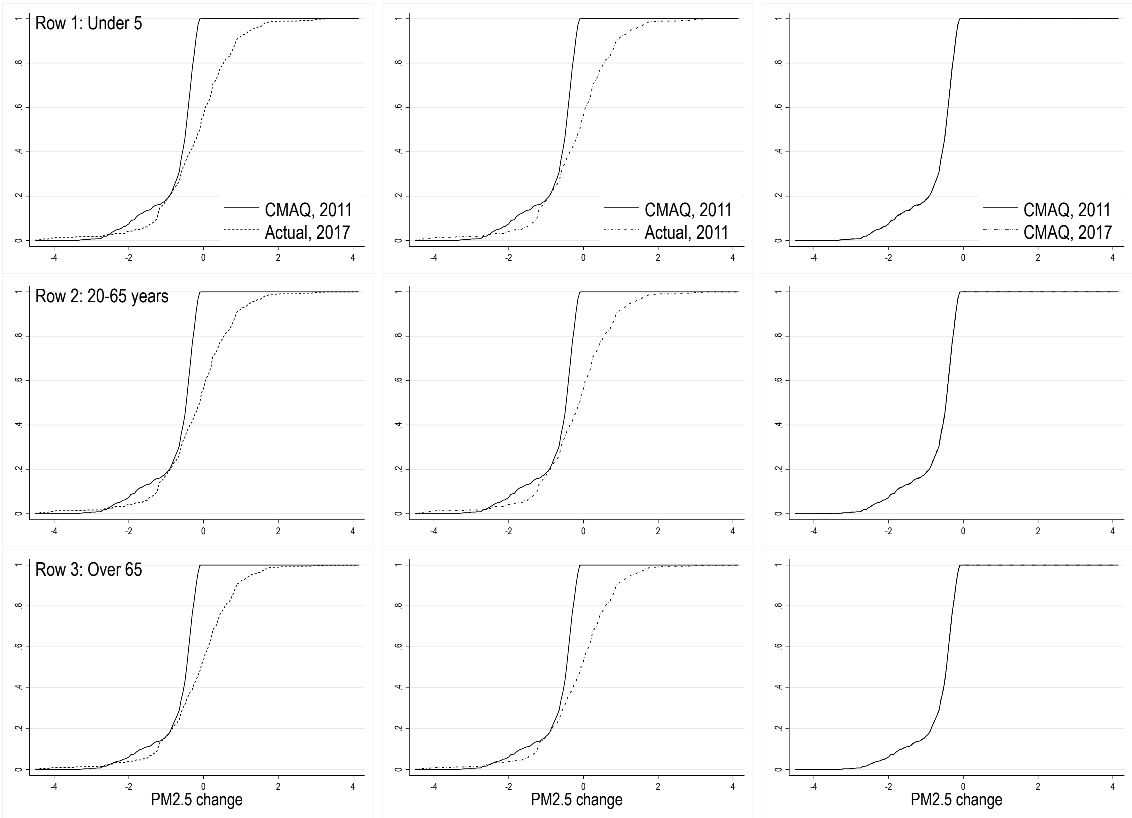
(a) *Overall error* (b) *Model error* (c) *Sorting error*



Note: Column (a) plots the fraction of the group population in 2011 with at least each level of CMAQ predicted change (after - before ECA) in PM2.5 or less (solid) and the fraction of the non-hispanic black population in 2017 with at least each level of actual change (after - before ECA) in PM2.5 or less (dashed). Column (b) plots again the fraction of the group population in 2011 with at least each level of CMAQ predicted change (after - before ECA) in PM2.5 or less (solid) and the fraction of the group population in 2011 with at least each level of actual change (after - before ECA) in PM2.5 or less (dashed). Column (c) plots again the fraction of the group population in 2011 with at least each level of CMAQ predicted change (after - before ECA) in PM2.5 or less (solid) and the fraction of the group population in 2017 with at least each level of CMAQ predicted change (after - before ECA) in PM2.5 or less (dashed). Groups include non-hispanic whites (Panel A), non-hispanic blacks (Panel B), hispanics (Panel C), and non-hispanic other (Panel D). The sample includes tracts with a balanced PM2.5 monitor within 50 km in counties with population-weighted centroid within 200 km of ship traffic.

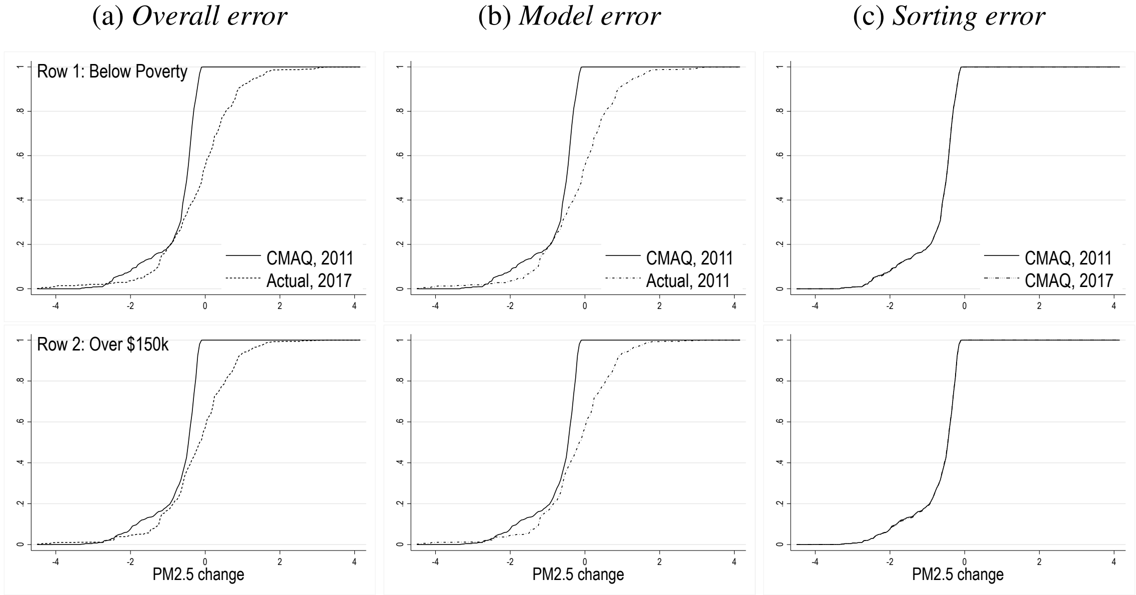
Figure 8: **Error Comparisons by Age**

(a) *Overall error* (b) *Model error* (c) *Sorting error*



Note: Columns (a)-(c) repeat Figure 7 Groups include individuals under 5 years old (Panel A), 20-65 years old (Panel B), and over 65 years old (Panel C). The sample includes tracts with a balanced PM2.5 monitor within 50 km in counties with population-weighted centroid within 200 km of ship traffic.

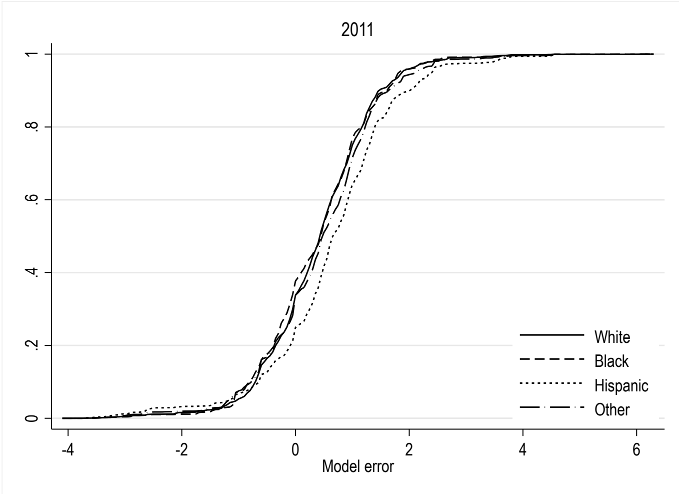
Figure 9: **Error Comparisons by Income**



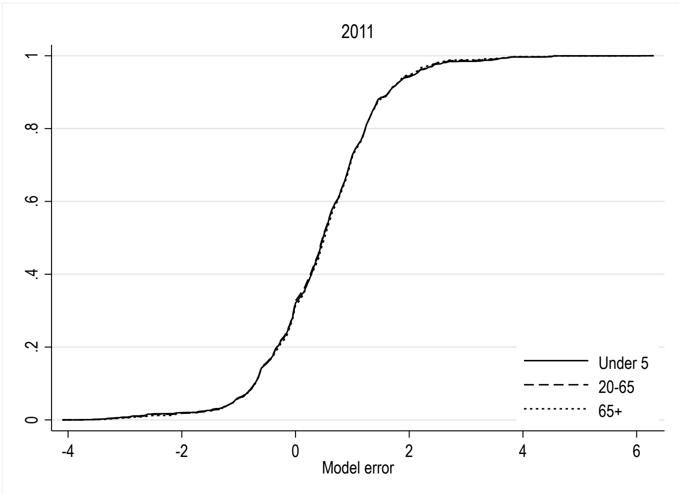
Note: Columns (a)-(c) repeat Figure 7. Groups include households below poverty (Panel A) and households with income above 150,000 USD (Panel B). The sample includes tracts with a balanced PM2.5 monitor within 50 km in counties with population-weighted centroid within 200 km of ship traffic.

Figure 10: **Model Error Comparisons Across Groups**

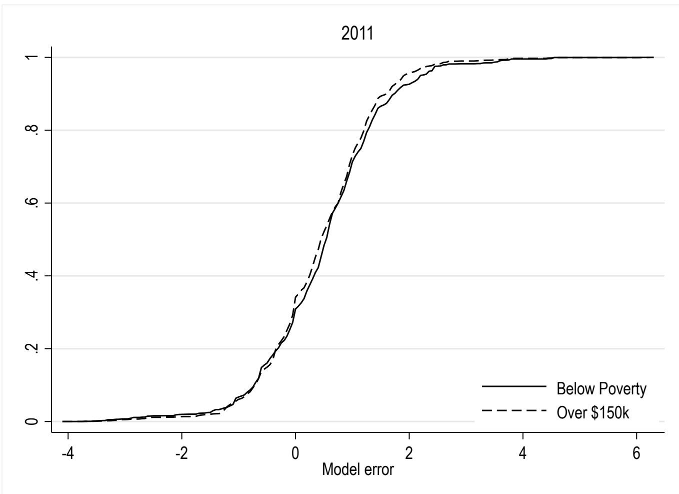
(a) *Racial Groups*



(b) *Age Groups*

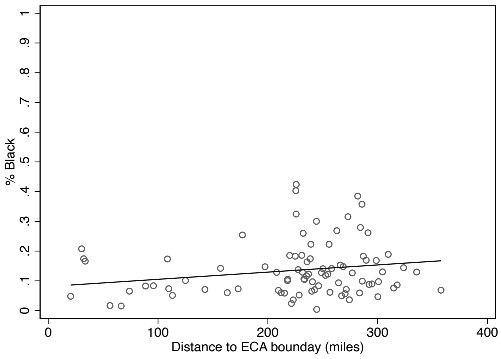


(c) *Income Groups*

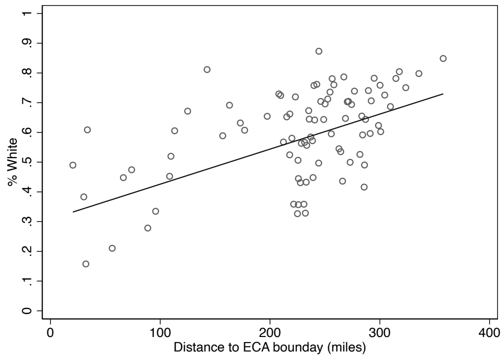


Note: The running variable, Model error, is the difference between actual PM2.5 change (after-before) and CMAQ predicted PM2.5 change (after-before). Positive values of Model error indicate actual PM2.5 declined after the ECA less than the CMAQ model predicted (e.g. CMAQ over-predicted the improvement) whereas negative values of Model error indicate actual PM2.5 declined after the ECA more than the CMAQ model predicted (e.g. CMAQ under-predicted improvement). The figure plots the fraction of the 2011 non-Hispanic white population (solid line), black (dash line), other (dash-dot line), and Hispanic (dot line) with at least each level of CMAQ error or less.

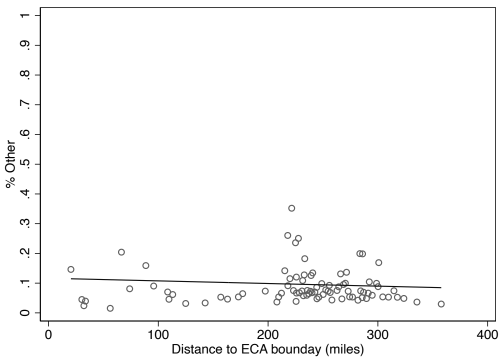
Figure 11: **Distance to ECA Boundary and Demographic Characteristics**



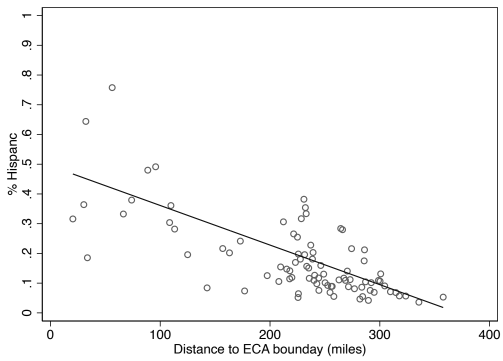
(a) Black non-hispanic



(b) White non-hispanic



(c) Other non-hispanic



(d) Hispanic

9 Tables

Table 1: **Balance of tracts with and without air quality monitor**

	(1) All in 200km	(2) In 200km and PM2.5 monitor
Race group:		
White	0.57	0.56
Black	0.15	0.14
Other	0.09	0.10
Hispanic	0.20	0.20
Age group:		
Under 5	0.06	0.06
20-65	0.61	0.61
65+	0.14	0.13
Income group:		
Below Poverty	0.11	0.11
Over $150k	0.11	0.12
Other characteristics:		
2011 Population	4,356.98	4,389.51
CMAQ Improvement	0.63	0.65
N Tracts	34,120.00	31,741.00

Note: The unit of observation in the 2011 census tract. The sample in column 1 includes all census tracts with in counties with population-weighted centroid within 200 km of heavy ship traffic. Column 2 drops tracts further than 50 km from a PM2.5 monitor with at least one observation per year from 2009-2016. Means are reported for the fraction of the 2011 population in each racial group, age group, and income group. The other reported characteristics include the average population in the tract, the average of the CMAQ PM2.5 improvement from the ECA in μgm^{-3}, and the total number of tracts.

Table 2: **Error Comparisons with Residual PM2.5**

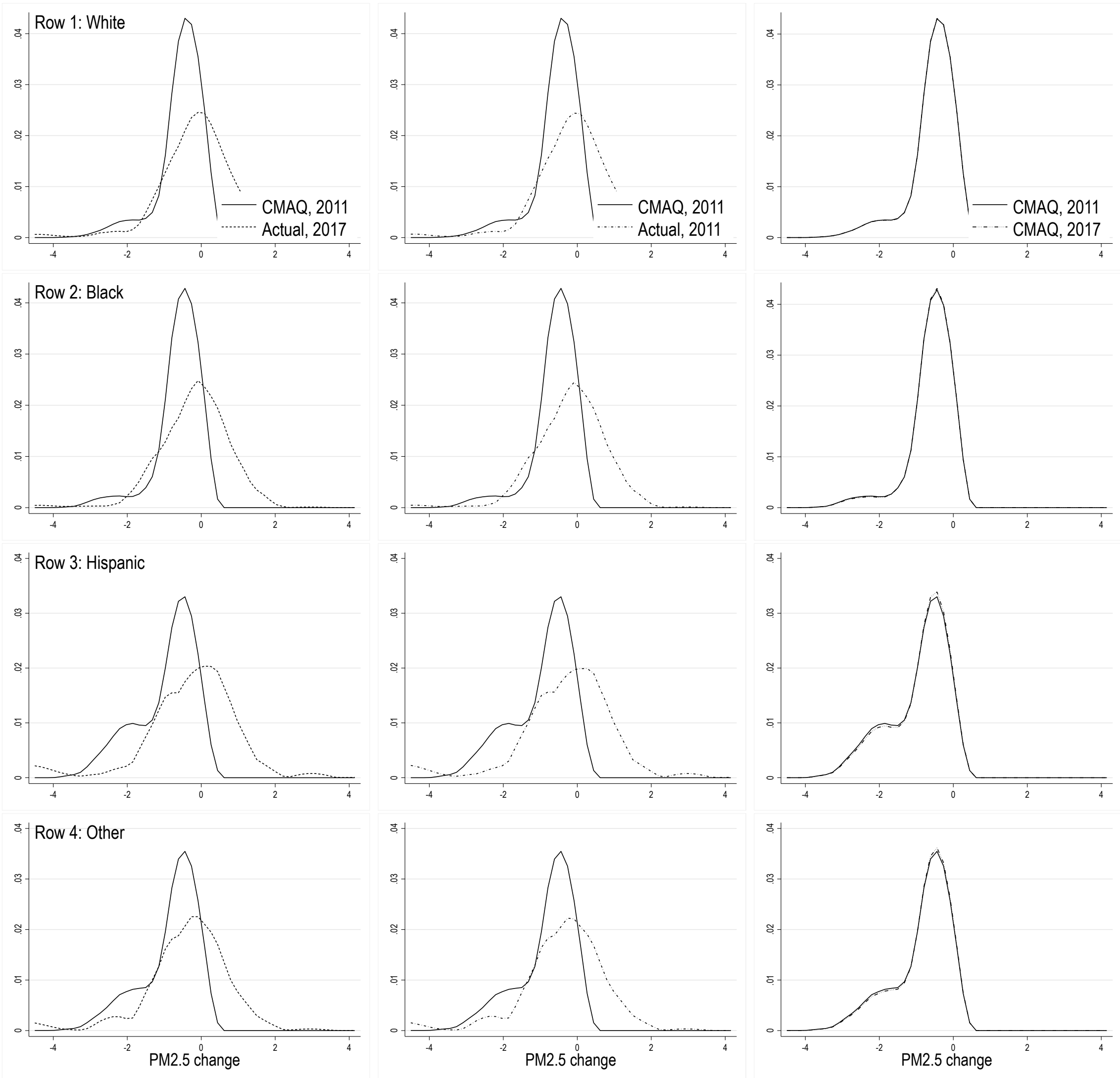
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Intended	Realized	Overall error		Model error		Sorting error	
	CMAQ/ 2011	PM25/ 2017	Difference (1)-(2)	PM25/ 2011	Difference (1)-(4)	CMAQ/ 2017	Difference (1)-(6)	LowSize/ 2011
Race group:								
Black	50.3	37.2	13.1	37.7	12.7	50.2	0.1	20.9
Hispanic	69.5	40.0	29.5	40.9	28.7	68.6	0.9	49.0
Other Race	63.2	45.2	18.0	45.8	17.4	62.3	0.9	28.6
White	43.2	33.0	10.2	33.0	10.2	43.4	-0.2	22.7
Age group:								
<5	53.5	37.2	16.3	37.1	16.4	53.8	-0.3	29.1
20-65	51.8	37.0	14.8	36.9	14.9	52.5	-0.7	28.2
>65	50.2	34.3	15.9	34.0	16.2	50.5	-0.4	30.3
Income group:								
Below Poverty	55.9	36.7	19.2	36.4	19.5	55.7	0.2	31.0
Above $150k	49.7	37.8	11.9	37.9	11.8	50.4	-0.7	23.0

Note: Column 1 reports the fraction of the 2011 population in each group that CMAQ predicted would obtain at least a 0.4 μgm^{-3} improvement in PM2.5. Column 2 reports the fraction of the 2017 population in each group that obtained at least a 0.4 μgm^{-3} improvement in actual PM2.5 Column 3 reports overall error, the difference between columns 1 and 2. Column 4 reports the fraction of the 2011 population in each group that obtained at least a 0.4 μgm^{-3} improvement in actual PM2.5. Column 5 reports model error, the difference between columns 1 and 4. Column 6 reports the fraction of the 2017 population in each group that CMAQ predicted would obtain at least a 0.4 μgm^{-3} improvement in PM2.5. Column 7 reports sorting error, the difference between columns 1 and 6. Column 8 reports the fraction of the 2011 population in each group located in tracts with less than 200 nautical miles from the exterior of the ECA.

10 Appendix

Figure A1: **Error Comparisons by Race: Probability Density Functions**

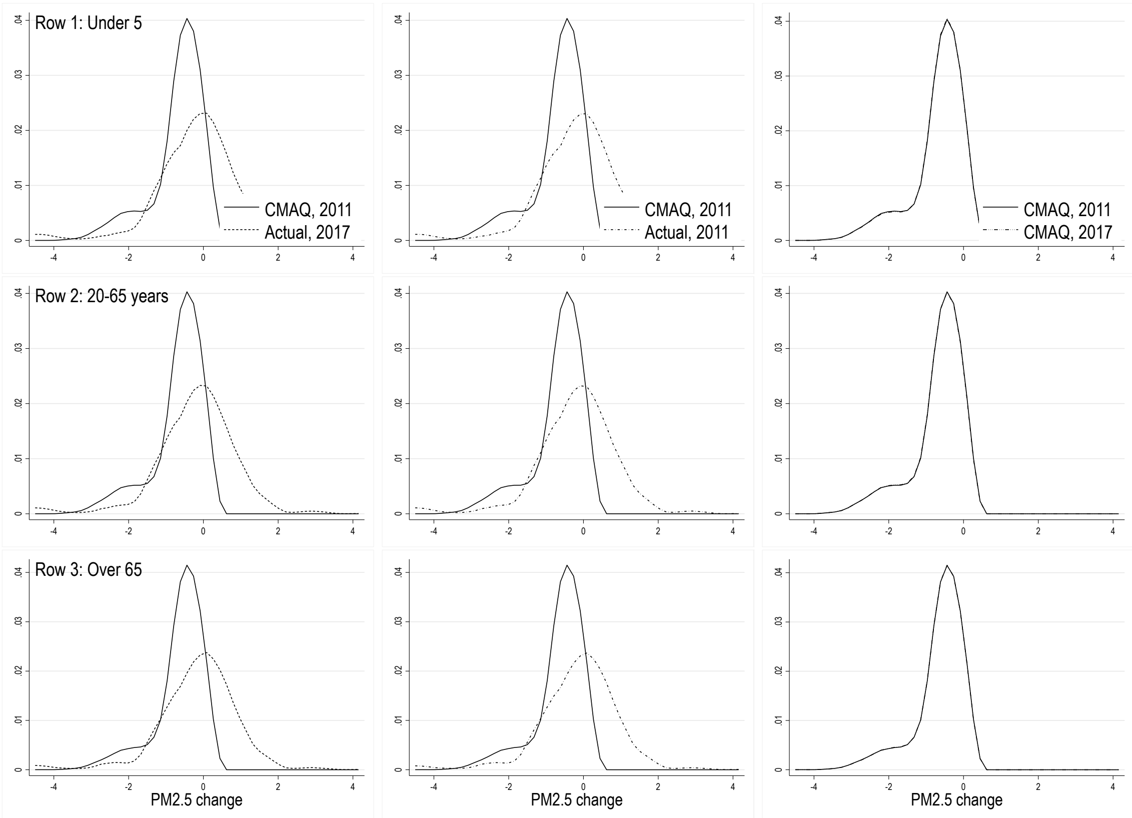
(a) *Overall error* (b) *Model error* (c) *Sorting error*



Note: Panel (a) plots the fraction of the group population in 2011 with each level of CMAQ predicted change in PM2.5 or less (solid) and the fraction of the group population in 2017 with each level of actual change (after - before ECA) in PM2.5 (dashed). Panel (b) plots again the fraction of the group population in 2011 with each level of CMAQ predicted change in PM2.5 or less (solid) and the fraction of the group population in 2011 with each level of actual change (after - before ECA) in PM2.5 (dashed). Panel (c) plots again the fraction of the group population in 2011 with each level of CMAQ predicted change in PM2.5 or less (solid) and the fraction of the group in 2017 with each level of CMAQ predicted change (after - before ECA) in PM2.5 (dashed). Groups include non-hispanic whites (Panel A), non-hispanic blacks (Panel B), hispanics (Panel C), and non-hispanic other (Panel D). The sample includes tracts with a balanced PM2.5 monitor within 50 km in counties with population-weighted centroid within 200 km of ship traffic.

Figure A2: **Error Comparisons by Age: Probability Density Functions**

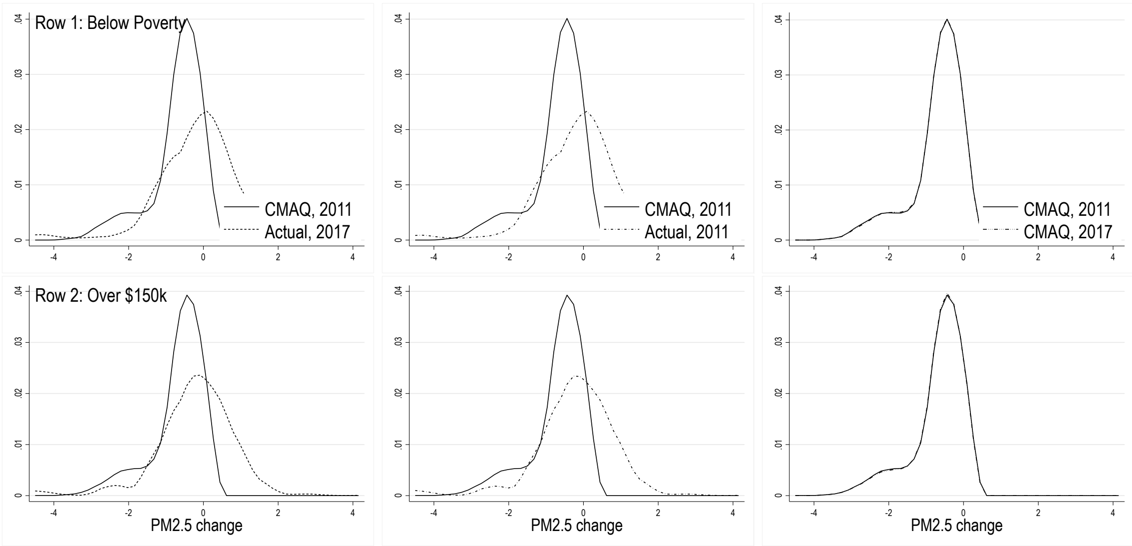
(a) *Overall error* (b) *Model error* (c) *Sorting error*



Note: Columns (a)-(c) repeat Figure A1. Groups include individuals under 5 years old (Panel A), 20-65 years old (Panel B), and over 65 years old (Panel C). The sample includes tracts with a balanced PM2.5 monitor within 50 km in counties with population-weighted centroid within 200 km of ship traffic.

Figure A3: **Error Comparisons by Income: Probability Density Functions**

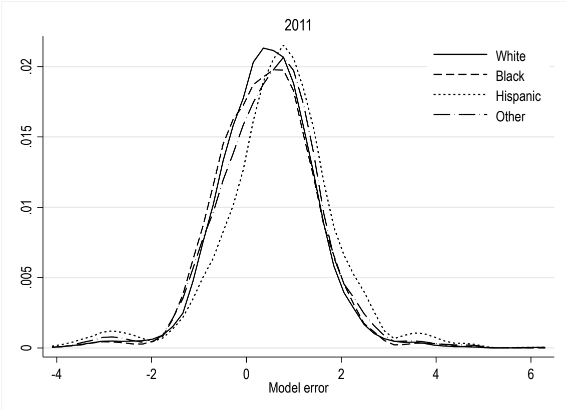
(a) *Overall error* (b) *Model error* (c) *Sorting error*



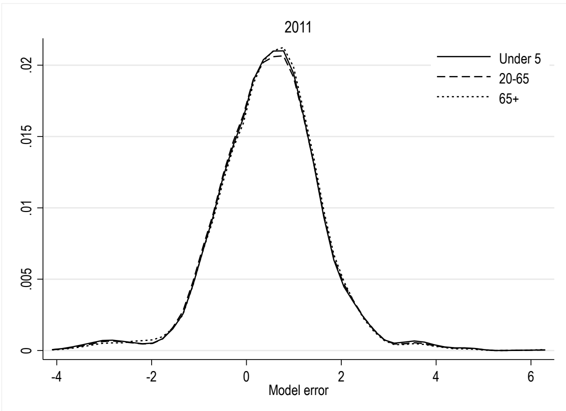
Note: Columns (a)-(c) repeat Figure A1. Groups include households below poverty (Panel A) and households with income above 150,000 USD (Panel B). The sample includes tracts with a balanced PM2.5 monitor within 50 km in counties with population-weighted centroid within 200 km of ship traffic.

Figure A4: **Model Error Comparisons Across Groups: Probability Density Functions**

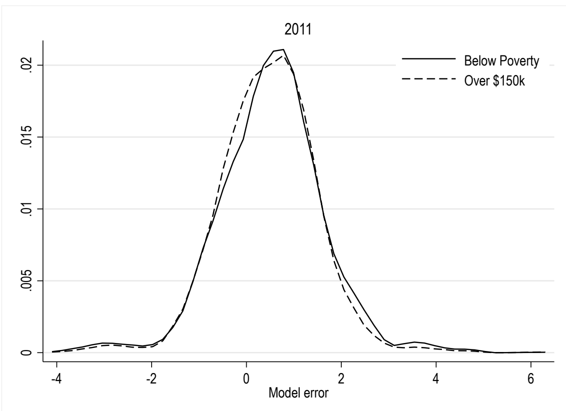
(a) Racial Groups



(b) Age Groups



(c) Income Groups



Note: The running variable, CMAQ error, is the difference between actual PM2.5 change (after-before) and CMAQ predicted PM2.5 change (after-before). Positive values of CMAQ error indicate actual PM2.5 declined after the ECA less than CMAQ predicted (eg CMAQ over-predicted improvement) whereas negative values of CMAQ error indicate actual PM2.5 declined after the ECA more than CMAQ predicted (eg CMAQ under-predicted improvement). Panel (a) plots the fraction of the 2011 population of each group at each level of CMAQ error. Panel (b) plots the fraction of the 2017 population of each group at each level of CMAQ error.

Table A1: Change in Demographic Characteristics by Census Tract

	(1)	(2)	(3)	(4)	(5)	(6)
	Change (2017-2011)			Change (2023-2017)		
	mean	sd	count	mean	sd	count
Race group:						
Black	-0.13	5.36	33723	-0.51	6.23	27540
Hispanic	1.62	6.29	33723	1.74	6.98	27540
Other Race	1.08	4.72	33723	2.34	5.48	27540
White	-2.57	7.16	33723	-3.57	7.98	27540
Age group:						
<5	-0.34	2.80	33723	-0.50	3.03	27540
20-65	-0.47	5.04	33723	-1.28	5.37	27540
>65	2.09	3.97	33723	2.15	4.54	27540
Income group:						
Below Poverty	0.32	7.33	33582	-1.96	7.92	27431
Above $150k	3.21	5.55	33608	10.66	8.16	27450

Table A2: **Extended Error Comparisons with Residual PM2.5**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Intended	Realized	Overall error		Model error		Sorting error	
	CMAQ/ 2011	PM25/ 2023	Difference (1)-(2)	PM25/ 2011	Difference (1)-(4)	CMAQ/ 2023	Difference (1)-(6)	LowSize/ 2011
Race group:								
Black	50.3	39.0	11.3	37.7	12.7	47.6	2.8	20.9
Hispanic	69.5	43.4	26.1	40.9	28.7	66.9	2.6	49.0
Other Race	63.2	46.1	17.1	45.8	17.4	60.0	3.2	28.6
White	43.2	34.6	8.6	33.0	10.2	42.0	1.2	22.7
Age group:								
<5	53.5	39.3	14.1	37.1	16.4	51.6	1.9	29.1
20-65	51.8	39.3	12.5	36.9	14.9	51.6	0.2	28.2
>65	50.2	36.9	13.2	34.0	16.2	50.4	-0.2	30.3
Income group:								
Below Poverty	55.8	38.1	17.6	36.0	19.7	54.3	1.5	32.4
Above $150k	49.7	40.0	9.7	37.9	11.8	50.2	-0.5	23.0

Note: Column 1 reports the fraction of the 2011 population in each group that CMAQ predicted would obtain at least a 0.4 μgm^{-3} improvement in PM2.5. Column 2 reports the fraction of the 2023 population in each group that obtained at least a 0.4 μgm^{-3} improvement in actual PM2.5. Column 3 reports overall error, the difference between columns 1 and 2. Column 4 reports the fraction of the 2011 population in each group that obtained at least a 0.4 μgm^{-3} improvement in actual PM2.5. Column 5 reports model error, the difference between columns 1 and 4. Column 6 reports the fraction of the 2023 population in each group that CMAQ predicted would obtain at least a 0.4 μgm^{-3} improvement in PM2.5. Column 7 reports sorting error, the difference between columns 1 and 6. Column 8 reports the fraction of the 2011 population in each group located in tracts with less than 200 nautical miles from the exterior of the ECA.