Summer 8-1-2020

Essays on Pollution, Health, and Education

Wes Austin
Georgia State University

ABSTRACT

ESSAYS ON POLLUTION, HEALTH, AND EDUCATION

By

WES AUSTIN

August 2020

Committee Chair: Dr. Tim Sass

Major Department: Department of Economics

Coal ash accounts for one third of industrial water pollution in the United States. In Chapter 1, I assess the relationship between coal ash surface water discharges and three relevant outcomes: surface water quality, municipal system water quality, and fetal health indicators from a birth certificate database in North Carolina. Identification relies on geographic variation in downstream status of monitoring sites and municipal water intake locations, plant closures or conversions, and the relative quantity of coal ash released over time. I find that coal ash releases are associated with higher conductivity and pH in both downstream surface waters and municipal water supplies sourced from these waters. Water systems affected by coal ash tend to have more Safe Drinking Water Act violations for disinfectant byproducts, inorganic chemicals, and health-based violations. I quantify the costs of coal ash water pollution with respect to fetal health and home sales. Exploiting variation arising from mothers' moves, I find that a newborn potentially exposed to coal ash water pollution is 1.7 percentage points more likely to have low birthweight compared to an unexposed sibling. I conclude by estimating how a legislative act mandating drinking well testing affected home sale prices in regions around coal ash plants. After the act, sale prices of homes within 1 mile of coal ash ponds declined by 12-14%, or over $37,000.

Chapter 2 investigates how school-age children are affected by diesel emissions from school buses. Diesel emissions from school buses expose children to high levels of air pollution; retrofitting bus engines can substantially reduce this exposure. Using variation from 2,656 retrofits across

Georgia, we estimate effects of emissions reductions on district-level health and academic achievement. We demonstrate positive effects on respiratory health, measured by a statewide test of aerobic capacity. Placebo tests on body mass index show no impact. We also find that retrofitting districts experience significant test score gains in English and smaller gains in math. Our results suggest that engine retrofits can have meaningful and cost-effective impacts on health and cognitive functioning.

Chapter 3 explores farm-to-school policies. School meal provision represents one of the largest food markets in the country. In 2015, 42,000 schools serving 23.6 million students engaged in farm-to-school nutrition sourcing policies. Yet, little is known about how much school systems actually source their food locally or about the average relationship between farm-to-school policy adoption and local sourcing of school food. I link 17 years of school district nutrition expenditures across the state of Georgia to a unique commodity-by-county survey of agricultural revenues to assess how much school systems source food from within their county and neighboring counties. I then incorporate four years of survey-based information on district farm-to-school policies to test how farm-to-school programs differentially impact local sourcing patterns. Identification comes from spatiotemporal variation in school district adoption of a farm-to-school policy and variation in expenditures associated with the community eligibility provision of the Healthy Hunger Free Kids Act. Results suggest that as much as $966M of school nutrition expenditures flow to producers within the same county. Of this, perhaps as much as $680M, or 0.6% of all agricultural revenues in the state from 2001-2017, are associated with adoption of farm-to-school policies by school districts.

ESSAYS ON POLLUTION, HEALTH, AND EDUCATION

BY

WES AUSTIN

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree
of
Doctor of Philosophy
in the
Andrew Young School of Policy Studies
of
Georgia State University

GEORGIA STATE UNIVERSITY
2020

Acknowledgments

This dissertation would not have been possible without generous support from countless individuals in academia, government, the private sector, and my home. Unable to list every person, I will restrict myself to specifically thanking my family members, Shannon Bryant, Miranda Burke, Luc Frolet, Jordan Brown, Nora Staikova, Mark Powell, George Davis, Siyu Pan, Jordan Jones, Elliott Rountree, and my dog Blaise. I give a special thank you to my advisers Tim Sass, Dan Kreisman, Garth Heutel, and Dajun Dai, among many other students and faculty at Georgia State University. In addition, data suppliers and seminar participants provided time and effort to helping this study. I thank each in the following paragraphs.

For chapter 1, I thank Matt Avery and the North Carolina State Center for Health Statistics for their support and provision of natality statistics. I am grateful to Jovian Sackett and the Southern Environmental Law Center for generously providing their water system intake database. I would also like to thank Eric Chai, Sean McGuire, Andrew Pitner, Julia Cavalier, Debra Watts, Eric Smith, and Nat Wilson with the North Carolina Department of Environmental Quality. Carla Foy, Clara Davis, Tamara Frank, Benny Laughlin, Mary Ann Fuller, and Robert Edelman with various state environmental protection divisions deserve additional credit. I thank workshop participants at Camp Resources XXVI and the 21st Colorado University Environmental and Resource Economics Workshop.

For chapter 2, I would like to thank Jonathan Smith, Ariell Zimran, and seminar participants at TEAM-Fest, the Southern Economics Association annual meeting, the Association of Education Finance and Policy annual conference, the Allied Social Science Associations (ASSA) annual conference, and the University of South Carolina for helpful comments. I am also grateful to Stacy Allman, Lou Ann Carmichael, and many others with the Environmental Protection Division of the Georgia Department of Natural Resources for their guidance on the Georgia school bus retrofit program. We are also grateful to Therese McGuire and Mike Tenoschok, with the Georgia Department of Education. I thank the GaDER program for assistance in identifying retrofits.

For chapter 3, I am grateful to Abbie King with Georgia Organics for kindly providing survey information on district farm-to-school programs, Karen Stubbs and Sharon Kane with the Center for Agribusiness & Economic Development at the University of Georgia, and Nicholas Warner with the Fiscal Research Center at the Andrew Young School of Policy Studies. I would also like to thank seminar participants at the Southern Economics Association and the Agricultural Economics Association annual conferences.

Contents

List of Figures

List of Tables

Chapter 1

Throwing the Baby out With the Ashwater? Coal Combustion Residuals, Water Quality, and Fetal Health

1 Introduction

Coal combustion residuals (CCRs) are the waste material from burning coal. Also known as coal ash, 110 million tons of CCRs are produced each year in the United States, of which 2.7 million tons are released into surface waters.[1] The remainder is primarily stored in wet landfills, while roughly one quarter is recycled.[2] Although surface-water discharges of coal ash effluent represent a small fraction of all coal ash produced, they account for one third of all industrial water pollution by toxicity and one half by mass.[3] No previous study has estimated how coal ash surface water discharges affect municipal water quality and human health.

Coal ash threatens water supplies because of the relative toxicity of constituent compounds, the quantity produced, and the quality of many confinement landfills. Heavy metals including arsenic,

[1] MacBride (2013), Gollakota et al. (2019). Globally, 750 million tons were produced in 2015, up from 500 million tons in 2005.

[2] See Yao et al. (2015), Gollakota et al. (2019) for reviews of alternative uses of coal combustion residuals.

[3] Bernhardt et al. (2016), Boyce and Ash (2016).

selenium, cadmium, chromium, lead, and mercury compose at least one third of coal ash.[4] Coal ash contains over four times as many heavy metals by mass as parent coal due to combustion of organic compounds.[5] Of the 63 steam-generating coal power plants incorporated in this paper, the average plant has seven containment landfills totaling 176 acres at an average depth of 50 feet.[6] Over 130 of these ponds were built before 1980, and at least 141 have no impermeable lining to protect groundwater.[7] Confinement and disposal practices increase the risk of surface-water and groundwater contamination. In a recent report, the EPA documented 149 damage litigation cases of coal ash impoundments affecting groundwater and 152 of coal ash affecting surface water.[8] Although municipal water providers filter most of the harmful compounds in coal ash, disinfectants used to treat the water interact with remaining CCRs to create harmful compounds known as disinfectant byproducts (DBPs). The formation of DBPs decreases the effectiveness of disinfectants.[9] Moreover, changes to the properties of water such as pH, temperature, and conductivity can affect corrosivity of pipes, leading to increased lead and copper levels in drinking water. While water quality in the developing world is known to affect human health, few studies have investigated how municipal water quality in a developed country may affect health.[10]

The purpose of this paper is to determine how CCR water pollution affects municipal water quality, human health, and home values. First, I replicate and generalize previous findings that CCRs affect surface water quality using a larger geographic region and longer time horizon. Next, I estimate the impact of this surface water pollution on measures of municipal water quality and the likelihood of a Safe Drinking Water Act Violation. Then, I assess whether this point-source pollutant may affect human health. Finally, I quantify residential willingness-to-pay to avoid coal ash water pollution. To answer these questions, I obtain six types of information: annual coal ash

[4]EPA (2015a), Shy (1979), Munawer (2018), Ibrahim (2015), Izquierdo and Querol (2012).

[5]Yao et al. (2015).

[6]Ash (2019) For comparison, Disney Land is 85 acres.

[7]Many inactive ponds lack information on construction date or lining status. See Table B.1 for summary statistics on coal ash containment facilities.

[8]EPA (2015a).

[9]EPA (2001), Davison et al. (2005), Wang et al. (2012).

[10]Among many others, He and Perloff (2016), Currie et al. (2017), Troesken (2008), Cutler and Miller (2005), Jalan and Ravallion (2003), Brainerd and Menon (2014) explore the relationship between water quality and human health in developing countries. Currie et al. (2013) and Marcus (2019) use samples in New Jersey and North Carolina.

surface water releases across 63 power plants, surface water monitoring tests over six states, municipal water quality monitoring tests over five states, municipal water quality violations over six states, birth certificates for 1.5 million children born in North Carolina, and home sale records for twelve counties in North Carolina. At a minimum, the sample covers 2005 to 2017. Identification of water quality changes associated with coal ash pollution relies on three forms of variation. The first form of variation is "downstream" status of monitoring sites or municipal water intakes within watershed regions. The second is temporal variation in the operating status of upstream coal ash facilities, which arises from plant closures and changes to confinement practices. The third is the relative quantity of coal ash released upstream from a water quality monitor or intake site, which occurs naturally over time and also due to plant closures and conversions. To test for health effects from water quality changes, I follow the literature in comparing fetal health indicators of siblings exposed to differential water quality.[11] I conclude by estimating household willingness-to-pay to avoid coal water pollution using repeat sales of homes near ash facilities in North Carolina. In this estimation procedure, I exploit a legislative change leading to the discovery of unsafe drinking water in many home wells surrounding coal ash plants. This study is the first to directly assess the impact of coal ash water pollution on drinking water supplies over a large geographic region and time horizon. I also add to a limited literature on the effect of water quality on fetal health outcomes in a developed-country context. Estimation of the willingness-to-pay to avoid contaminated drinking wells has broad relevance to both the housing value effects of environmental crises and risk perception among households near potential disaster sites.[12]

I find that contemporaneous coal ash releases increase the concentration of heavy metals in downstream surface waters; these include arsenic, lead, and selenium. Surface water quality monitors downstream from coal ash release sites also have altered properties. They tend to have higher conductivity, lower dissolved oxygen, higher pH, and higher temperature. Municipal water systems sourcing from waters potentially affected are also more likely to have higher conductivity, an indicator of elevated suspended and dissolved compounds. These water systems experience more

[11] Currie et al. (2013).

[12] Christensen et al. (2019b), Coulomb and Zylberberg (2016).

water quality violations for disinfectant byproducts, inorganic compounds, arsenic, and maximum contaminant level violations. I find evidence that maximum contaminant level, reporting, inorganic compound, arsenic, and health-based violations are driven by contemporaneous releases of coal ash pollution. Children born in residences served by municipal water systems downstream from active coal ash sites, in comparison to unexposed siblings, are 1.7 percentage points more likely to have birthweight. On average, these newborns weigh 1.2 ounces less than unexposed siblings, and they're 1.2 percentage points more likely to be preterm. Newborns of mothers with less education are more affected by coal ash pollution than average exposed children. These effects are driven both by adverse outcomes of mothers moving into coal ash municipal water service zones and by improvements for mothers moving out of coal ash municipal water service zones. Finally, residences within 1 mile of a coal ash pond, after discovery of well water considered unsafe to drink by the EPA, sold for $37,000 - $45,000 less than previously. Results provide strong evidence that coal ash water pollution negatively affects surface water quality and complicates the municipal water treatment process. These changes to municipal water quality likely affect human health, and the analysis of repeated home sales reveals that households care greatly about potential exposure to this form of pollution.

2 Motivation, Prior Work, and Contribution

An extensive literature documents the negative health consequences of exposure to coal through kitchen handling, home heating, mine drainage, mining dust, shipping and stockpile dust, and smokestack emissions.[13] These health consequences are large both in magnitude and relative to the cost of the coal.[14] Only one study investigates the health effects of coal ash water contamination. The study found that coal-polluted well water is associated with skin cancers, toxicities to internal

[13]Liu et al. (2002), Barreca et al. (2014), Kravchenko and Lyerly (2018), Pershagen et al. (1986), Clay et al. (2015, 2016).

[14]Jha and Muller (2017) found that the external costs from coal stockpile dust were four times the per-ton cost of the coal itself.

organs, neuropathy, nephrotoxicity, cirrhosis, ascites, and liver cancer.[15] However, the study relates to household disposal of cooking coal ash near shallow drinking wells rather than industrial coal ash containment practices, and it is also set in a developing country. In a recent literature review on the health effects of coal combustion residuals from steam power plants, the author found no study quantifying the extent of drinking water quality concerns and recommended future studies on the range of individual exposures to coal ash contaminants from water.[16]

CCRs primarily affect surface and ground waters in three ways. First, ash ponds are occasionally or continually drained into nearby bodies of water. CCRs also seep through the sides of containment facilities. Because coal plants and ash ponds are constructed next to large bodies of water, seepage is nontrivial.[17] Third, pressure from the weight of additional CCRs and water cause a leachate of dissolved compounds to flow into groundwater if a containment pond is unlined or poorly lined, affecting public and private wells and eventually also surface waters.[18] A broad literature demonstrates the chemical profile of coal ash water pollution, the conditions under which coal ash is mobilized, and the characteristics of affected surface waters.[19] In general, these studies cover relatively small geographic regions and provide a snapshot temporal view of local water quality.[20]

CCR source-water contamination may affect drinking water quality through the formation of disinfectant byproducts, corrosion of pipes, and residual contaminants after water treatment. Coal ash effluent increases the quantity of total dissolved solids in drinking water supplies, which is associated with increased formation of trihalomethanes, a group of disinfectant byproducts, during water treatment.[21] Bromide, a relatively harmless constituent of coal ash, interacts with chlorine

[15] Yu et al. (2007).

[16] Kravchenko and Lyerly (2018).

[17] Coutant et al. (1978) compare intentional water discharges with seepage water, finding that the latter contained 44 times the amount of dissolved iron and had a pH of 2.9; both sources of water killed all experimental fish subjects within 72 hours, with the seepage water killing all fish within the first 24. Unexposed control fish populations experienced no mortality.

[18] Of 14 North Carolina large coal ash confinement facilities, two thirds leach pollution into groundwater.

[19] Kopsick and Angino (1981), Baba and Kaya (2004), L. Carlson and C. Adriano (2009), EPA (2015a).

[20] An exception is EPA (2015a), which creates a model to estimate the effect of coal ash effluent discharges on nearby surface waters, taking characteristics of the pond and nearby body of water into consideration. The study examines five sites across the country, and uses the analysis to make effluent limitation policy suggestions.

[21] Handke (2009).

to form another group of disinfectant byproducts, haloacetic acids.[22] Corrosivity is the rate of pipe oxidation; high corrosivity indicates the potential for leaching of pipe materials such as lead and copper into drinking water. PH, conductivity, total dissolved solids, alkalinity, temperature, dissolved oxygen, and total hardness influence the corrosivity of water. Corrosivity is a major health concern for untreated groundwater sources.[23] However, fluctuations in surface water quality leading to corrosivity changes may also pose a public health concern.[24] For example, chloride in coal ash, increasingly present in US surface waters, affects corrosivity and hence lead levels in drinking water.[25] Properties of water related to coal ash, such as pH, also affect corrosivity and may impact human health. Clay et al. (2010) take advantage of variation in pipe materials and water pH across regions of the US from 1900-1920, finding that a slight normalizing of pH in locations with lead pipes would decrease fetal mortality by 7-33%. Troesken (2008) finds a similarly strong relationship between pH, lead pipes, and fetal health. Finally, variations in pollution releases, weather events, and accidents may impact the efficacy of treatment systems designed for different source-water quality.[26]

Animal-based studies demonstrate that coal ash water pollution harms the reproductive health of many organisms.[27] The potential influence of coal ash water pollution on pipe corrosion may also signal a public health concern because lead impairs child and fetal development.[28] Further, disinfectant byproducts may affect fetal health even if the same compounds in similar doses are low-risk to adults.[29] Prior work causally associates differential water quality with an increased risk of low-birthweight newborns, providing a basis for investigating whether residual coal ash pollutants, materials from pipe corrosion, or disinfectant byproducts may impact fetal health.[30] I use fetal health indicators for this analysis because of the greater vulnerability of newborns to pol-

[22]Heller-Grossman et al. (1993), Cowman and Singer (1996), Liang and C Singer (2003).

[23]One third of drinking water wells in the United States have potentially corrosive water (Belitz et al., 2016).

[24]Singley et al. (1984), Neffand et al. (1987).

[25]Zhu et al. (2008), Stets et al. (2012).

[26]Davison et al. (2005).

[27]Gillespie and Baumann (1986), Heinz and Hoffman (1998), Hopkins et al. (2002).

[28]Gazze (2015), Clay et al. (2010, 2018, 2019), Miranda et al. (2007).

[29]Studies suggest that DBPs increase risk of bladder cancer when ingested at levels currently observed in industrialized countries (Cantor et al., 2010, Villanueva et al., 2004).

[30]Currie et al. (2013).

lution. Low-birthweight newborns are also costly to society. They are more prone to chronic and degenerative conditions like diabetes and heart disease; they also have lower test scores, educational attainment, and income.[31] The short time window of gestation also increases the likelihood of noticing health impacts that would take longer to manifest in adults and likely coexist with many pollutant exposures.

This study contributes to several literatures. I generalize previous findings on the effects of coal disposal practices on surface water quality to a region of six states, thirteen years of monitoring tests, and a wide array of compounds. I also contribute to a limited literature on the role that point-source pollution plays on municipal water quality, providing a relatively novel outcome in the form of regular state monitoring tests. In so doing, I provide the first evidence on the contemporaneous relationship between coal ash water pollution and municipal water quality. Adding to other studies on the fetal health consequences of local pollution, I estimate the relationship between coal ash sites and indicators of fetal health, incorporating both air and water quality information.[32] This study adds to our understanding of the life-cycle costs of coal, as many papers disregard water quality costs except those related to mining.[33] Similarly, the study provides an additional context through which to view the benefits of surface-water pollution abatement, recently found to be less than one fourth the costs of cleanup grants in Keiser and Shapiro (2017). Indeed, the EPA's own analyses rarely find that water quality regulations pass a cost-benefit test, with a median benefit-cost ratio of 0.37 across all regulations over the past several decades according to a recent study.[34] Because these previous cost-benefit analyses do not include health benefits via the municipal drinking water mechanism, this study sheds light on how a potentially missing benefit may affect the results of decades of federal cost-benefit analysis on surface water quality regulations.

[31] Osmond and Barker (1991), Almond and Currie (2011).
[32] Currie and Walker (2011), Currie et al. (2017), Persico et al. (2016), Jha and Muller (2017).
[33] Amigues et al. (2011), Muller et al. (2011).
[34] Keiser et al. (2019).

3 Data

This study incorporates information on coal ash disposal practices, surface water quality, municipal water quality, natality outcomes, air pollution, and home sales. In the following sections, I summarize average differences across potentially affected and likely unaffected surface waters, municipal water systems, newborns, and homes. For detailed description of how I geographically assign treatment indicators, see chapter appendix sections 8.1, 8.2, and 8.3.

3.1 The Quantity and Location of Coal Ash Disposal

The Toxic Releases Inventory (TRI) provides facility-by-year-by-pollutant information on the quantity of over 650 regulated substances released into the environment. Many of the compounds present in coal ash are regulated substances. All facilities releasing at least one of these compounds and employing at minimum ten employees must report their pollution release information annually, ensuring that industrial steam-generating coal power plants are included in the TRI.[35] The pollutant compounds are split up by type of release, allowing separation of the quantity that is released into surface waters from the quantity that is impounded. I combine TRI reports with information on the age, depth, and lining status of each plant's confinement ponds or landfills assembled by the non-profit Southeast Coal Ash. I limit my sample of coal plant release sites to those with positive water pollution from 2005 to 2017 across six southern states. These states are Alabama, Georgia, North Carolina, South Carolina, Tennessee, and Virginia. Power plants not combusting coal were excluded from the sample. The remaining sample includes 63 steam electricity generating coal power plants. These sites are mapped in Figure A.1. Table B.1 displays annual facility-level information on coal ash loadings from 2005-2017, including toxicity weights for many of the constituent compounds of coal ash.[36] Additionally, Figure A.4 shows the annual

[35]Self-reporting allows the possibility of under-reporting and measurement error. To the extent that firms under-report true pollution releases, regression estimates would be biased to zero. To limit the influence of mis-measured or poorly-estimated release figures by pollutant, I employ models with a binary indicator for whether surface-water pollution releases occurred and others with a variable for the total coal ash surface-water releases across all compounds.

[36]The EPA's Risk-Screening Environmental Indicators (RSEI) toxicity weights allow comparison of the toxicity of different compounds compiled in the TRI. See https://www.epa.gov/rsei/rsei-toxicity-weights for more information.

average distribution of toxic releases of coal facilities. The same figure plots these average releases over time. The average coal power plant releases approximately 10 tons of coal ash compounds into surface waters each year, and this level has remained roughly constant over the sample excluding a disastrous 2008 spill at the Kingston Fossil Plant in Tennessee. However, this average masks significant heterogeneity in the quantity of surface water pollution across plants. Meanwhile, the quantity of coal ash impounded in confinement landfills has decreased over the past decade from around 400 tons per plant per year to a little over 250 tons. Figure A.5 provides a breakdown of coal ash surface water loadings by type of chemical. Of the tons that are released into surface waters, the bulk of the pollution is composed of relatively harmless compounds such as barium, copper, manganese, and nickel. However, it is not uncommon for plants to release multiple tons of more harmful compounds such as arsenic, chromium, lead, and vanadium into nearby surface waters in any given year.

3.2 Surface Water Quality Monitoring Information

I retrieve surface water quality information from the Water Quality Portal (WQP), the largest standardized water quality dataset currently in existence.[37] Developed by researchers from the U.S. Geological Survey, the Environmental Protection Agency, and the National Water Quality Monitoring Council, the WQP combines the USGS National Water Inventory System, USGS BioData, USDA Stewards, and EPA Storets databases. The WQP features 2.4 monitoring sites and roughly 300 million analyte results over many decades and thousands of compounds. Decisions underlying the location of monitors and timing of tests are not observable.[38] I limit the sample to monitoring sites located in lakes, rivers, and streams. I also limit the analysis to eight core water quality analytes known to be associated with coal ash water pollution; these include arsenic, chromium,

[37] Read et al. (2017). I use the DataRetrieval package in R to download and import the data (De Cicco et al., 2018).

[38] USGS hydrologists designed intentionally representative samples of US waters for common analytes such as pH and conductivity, but local governmental agencies and other researchers contributing to the WQP may have selected locations based on un-observable factors (Keiser and Shapiro, 2018). To limit the influence of selection, only monitors with at least three tests for a given compound are incorporated in regression models. See Figure A.2 for all monitor locations used in this paper.

conductivity, dissolved oxygen, lead, pH, selenium, and temperature.[39] See Table B.11 for a full list of compounds retrieved. All sample results that do not detect the tested compound are replaced with zeros, and I initialize an undetected flag for these observations. Measurements are converted to a standardized unit where possible (for example, milligrams/liter). Observations without convertible units of measurement are dropped.[40] After cleaning, the sample consists of 5.5 million measurements across 124,000 monitoring sites. Summary statistics are presented in Table B.2. I also compare how these water quality indicators change over time for monitors that are within 25 miles downstream of coal ash release sites in Figure A.6 and Figure A.7. For ease of visualization, I drop extreme outliers above the 99th percentile and non-standard samples before generating mean analyte levels over time.[41] The figures nevertheless generally confirm the summary statistics presented in Table B.4; coal ash affected waters have higher conductivity, pH, and temperature across the entire sample window. Dissolved oxygen levels are also often lower in affected regions than in unaffected regions. Affected regions tend to have lower average levels of common pollutants including lead, arsenic, selenium, and chromium, although the trends are noisy and include a spike in all compounds from 2008-2011 that may relate to differential testing priorities over time.

3.3 Municipal Water Quality Violations, Infrastructure, and Monitoring

The Safe Drinking Water Inventory System (SDWIS) houses municipal water system violation histories, water system summaries, water system details, and service zone geographic area. Violation history reports show when a water quality violation occurred, the nature of the violation, and the remediation action taken. Reports on water system summary, detail, and geographic area describe the population served, the number of facilities and service connections, and the population served by the water system.[42] Summary statistics for SDWIS violations are presented in Table B.4. Water systems affected by coal ash tend to be much larger and somewhat older than unaffected systems.

[39] EPA (2015a), Munawer (2018), Ibrahim (2015), Izquierdo and Querol (2012).

[40] I except pH from this decision rule and instead drop any pH observations outside the standard scale from 0-14.

[41] Standard refers to samples of surface water. Samples of sediments and hyporheic zones, which typically have different properties, are excluded from the figures but included in summary statistic tables and surface water regressions.

[42] Geographic service regions may be a town, a zipcode, or a county centroid if missing more precise information.

They also have more health-based Safe Drinking Water Act violations. Much of this average difference is evidently driven by violations for exceeding the maximum contaminant level of a given pollutant or breaking rules for arsenic, disinfectant byproducts, and inorganic chemicals. In Figure A.8, I plot the violation rate over time for affected water systems across six types of infraction. Affected water systems tend to have more maximum contaminant level and health-based violations over the entire sample window. In Figure A.9, I show that these infractions are primarily for breaking rules for inorganic chemicals and disinfectant byproducts. Water systems affected by coal ash tend to have lower violation rates for nitrates and coliform than unaffected systems. In Figure A.10, I break down all SDWA violations by type of infraction and state. Clearly, most of the maximum contaminant level violations relate to elevated levels of disinfectant byproducts. Violations for inorganic chemicals and volatile organic chemicals are primarily monitoring-based, which means that these systems are not testing for all required compounds. Finally, in Figure A.11, I show how these violations have trended over time by type of infraction and state. Monitoring violations appear to be the most common infraction type, and North Carolina tends to have the most SDWA violations since 2000.

I supplement SDWIS with state-provided water quality monitoring tests in Alabama, Georgia, North Carolina, South Carolina, and Virginia from 2005-2017.[43] These monitoring tests are used to determine violations of the Safe Drinking Water Act. Monitoring tests are samples of a water quality analyte taken at one facility.[44] According to the Safe Drinking Water Act, these monitoring tests must be performed by a third party at a frequency determined by the chemical and the population served by the water system.[45] 166 analytes are regularly tested across the sample states. These analytes may be grouped into 14 pollution classes. For all samples that do not detect the given compound, I replace the observed value with zero and initialize a non-detected flag. I also

[43]State agencies include the Alabama Department of Environmental Management, the Georgia Environmental Protection Division, the North Carolina Department of Environmental Quality, the South Carolina Department of Health and Environmental Control, the Tennessee Department of Environment and Conservation, and the Virginia Department of Environmental Quality. With the exception of Tennessee, each agency provided all available testing records over the sample window.

[44]For example, one observation may show that the level of lead in the water at a given facility on a given date was 0.005 mg/L.

[45]Currie et al. (2013).

generate indicators for the type of facility where a test occurred, allowing me to control for likely differences that may exist across tests at wells or intakes from those at treatment and distribution centers. Summary statistics for state-level monitoring tests are presented in Table B.4. Conductivity and arsenic, which are much higher in groundwater than in likely-affected surface waters, tend to be lower in affected municipal water systems. However, they generally have higher levels of disinfectant byproducts and pH. I show how these water quality indicators trend over time in Figure A.12. Across the entire sample window, affected water systems have higher levels of disinfectant byproducts.

I combine SDWIS data with state monitoring tests for two reasons. First, violation history provides a snapshot of municipal water quality. Samples conducted over time allow detection of more subtle differences in water quality that do not result in a violation. Second, the violation rate is an endogenous manipulable outcome.[46] It is likely that water systems sourcing from coal ash affected waters take precautionary treatment measures or perform compliance activities after any violation.

3.4 Birth Certificates and Fetal Health

The North Carolina State Center for Health Statistics provides birth certificate information for the period 2005-2017. These data report indicators of fetal health such as gestation length, birthweight, estimated gestation length, and presence of a congenital anomaly. They also include maternity characteristics such as age, education level, race, marital status, and smoking behavior.[47] The birth certificates track information on mother risk factors during pregnancy and delivery, such as hypertension, previous pregnancy termination, and number of prenatal visits. I obtained confidential records reporting mother's name and address at time of birth. Mother's full name, race, and birthday are used to link siblings. Mother's address of residence allows linking birth records

[46] Bennear et al. (2009).

[47] Paternal characteristics are limited to demographic information, and these records are often incomplete.

to specific water service regions.[48] Birth records with missing addresses or mother's names are excluded from the sample. Similarly, addresses not corresponding to a service zone are dropped from all regressions. A key difficulty in working with the natality statistics relates to the different birth certificate forms used over the sample period. Three types of reporting forms are used over the sample period; one covers 2005-2009, another covers the transition year 2010, and then a third is used for 2011-2017. Although all forms record certain information in the same format, such as birthweight and gestation length, other variables change across birth reporting forms. For example, race and education status report different categories across the two main reporting forms. Where possible, these measurements are adjusted to create temporally-consistent variables. Notably, congenital anomaly indicators cannot be properly conformed across the different forms due to certain conditions not being listed in the post-2010 form. This discrepancy results in apparently dramatically different rates of congenital anomalies in the pre-2010 and post-2010 forms.[49]

In Table B.3, I document systematic differences in fetal health across affected and unaffected mothers. Mothers ever served by municipal water systems affected by coal ash tend to have lower birthweight newborns and higher likelihood of preterm gestation. Affected mothers are more likely to be minority, unmarried, and have hypertension, although both affected and unaffected mothers tend to engage in similar rates of tobacco use and prenatal visits.[50] Interestingly, affected mothers are 5 percentage points more likely to move between pregnancies, perhaps reflecting perceived risk of coal ash pollution. Newborns of affected mothers are 0.8 ounces lighter, on average, and 0.5 percentage points more likely to have low birthweight (i.e., weigh less than 2500 grams). They also appear more likely to have congenital anomalies, although this discrepancy may relate to changes

[48]Property parcels, obtained from the NCSU GIS Library, were merged by spatial location using geographic coordinates and service zone polygons obtained from NC OneMap Geospatial Portal. Mother residence addresses were then merged to property parcels, and hence water service zones, using address, zipcode, and county names by a fuzzy-string matching algorithm, the stata package matchit (Raffo, 2015). Poor-quality matches were manually cleaned. Remaining unmatched addresses were assigned to water systems based on city of residence if the city is known to use coal ash affected water according to the Southern Environmental Law Center. See Appendix Section 8.3 for a lengthier description of the address matching procedure.

[49]For the purposes of this study, I exclude chromosomal anomalies and trisomy 21 from my indicator for a congenital anomaly because these conditions occur naturally in the human population and are not necessarily linked to pollution exposure.

[50]Lead exposure is associated with increased risk of hypertension (Gambelunghe et al., 2016).

in recording practices for this outcome. Figure A.13 displays four fetal health indicators over time between mothers ever potentially affected by coal ash and mothers likely not affected by coal ash.

3.5 Satellite-Based Monthly Air Quality

Air quality is an important determinant of fetal health.[51] I therefore incorporate satellite-based monthly fine particulate matter (i.e., particulate matter of size less than 2.5 micrometers in diameter) estimates as controls in the analysis. The Atmospheric Composition Analysis Group at Dalhousie University created these data by applying a machine-learning algorithm to repeated daily satellite images of aerosol optical depth, a measure of cloudiness, across small pixels on the earth's surface.[52] Using the extract raster to polygon feature in GIS software, I converted these pixel datapoints to county-level variables for the average, minimum, and maximum fine particulate matter for each month from 2000 to 2017. Infants are assigned air quality measurements based on the average and maximum county-level PM 2.5 reading over all months *in utero*. The advantage of satellite-based data is a wider coverage region than would be possible using air quality monitors, although prediction errors render these estimates less accurate for tiny regions or high pollution levels.[53] A recent study nevertheless demonstrates very similar fetal health outcomes using both satellite-based and monitor-based air quality measurements at the county level.[54]

3.6 Home Sale Prices

I obtain home sale tax records for twelve counties with coal ash ponds in North Carolina.[55] These records were obtained from multiple sources. County tax assessors provided property sales information for six counties. For another six counties, I purchased home sale information from CoreLogic's Configurable Real Estate Data Report. I merge each home address to a North Car-

[51] Chay and Greenstone (2003), Currie et al. (2008), Currie and Walker (2011), Jha and Muller (2017)

[52] van Donkelaar et al. (2019).

[53] Fowlie et al. (2019).

[54] Alexander and Schwandt (2019).

[55] Buncombe, Cleveland, Catawba, Chatham, Gaston, New Hanover, Person, Robeson, Rowan, Rockingham, Rutherford, and Stokes counties are included in the analysis.

olina property parcel database to extract geographic coordinates for all homes. I then use ArcGIS to merge these homes to a series of buffer polygons created around coal ash ponds at distances of 1, 2.5, and 5 miles. Because of the fragmented home sale source data, variables commonly used in hedonic housing analyses are mostly missing and often incongruous across counties. The exception is lot size. Summary statistics for home sales are presented in Table B.3. Figure A.14 plots average sale prices over time along with the distribution of sale prices in homes within 5 miles of a coal ash plant. Surprisingly, homes within five miles of a coal ash release site tend to be more expensive than more-distant homes over the entire sample period. On average, they sell for nearly $30,000 more than homes at greater distance from coal plants. They also have slightly more bedrooms and bathrooms than other homes in the same county, although their lot size is much smaller. These features of the data suggest that communities around coal ash ponds in North Carolina are more neighborhood-based and less rural than average properties in affected counties. Their proximity to large lakes and other recreation zones may also contribute to their higher sales prices.

4 Empirical Strategy

In the following sections, I describe the methods used to test the relationship between coal ash water pollution and surface water quality, municipal water quality, and fetal health. I also estimate how the revelation of unsafe well water affected home sale prices after a legislative act.

4.1 Surface Water Quality

I detect variations in surface water quality associated with coal ash water pollution with a surface water monitor fixed effects estimation procedure. Consider the following regression equation:

$$Y_{imwt} = \beta Ash_{it} + X_{it}\gamma' + \eta_i + \eta_{wm} + \eta_{wt} + \varepsilon_{imwt} \tag{1.1}$$

In Equation 1.1, Y_{imwt} is the level of arsenic, chromium, conductivity, dissolved oxygen, lead, pH, selenium, or temperature detected at a given monitor i in month m, watershed w, and year t.[56] Equation 1.1 includes fixed effects for monitor η_i, watershed-month η_{wm}, and watershed-year η_{wt}. X_{it} includes dummy indicators for sample medium type, a dummy indicator for abnormal weather event, dummy indicators for hydrologic condition type, and a dummy indicator if the analyte was not detected.[57] I two-way cluster all standard errors at the monitor and watershed level. I employ three versions of Ash_{it} to test related but distinct research questions. In the first, Ash_{it} is a time-invariant binary indicator equal to one if a monitor is downstream and within 25 miles of a release site.[58] In the second, Ash_{it} is a time-varying binary indicating whether upstream coal sites within 25 miles are actively releasing water pollution in year t according to the TRI.[59] In the third formulation, Ash_{it} is the annual quantity of coal ash released at all coal facilities within 25 miles upstream of monitor i.[60] For the monitor-constant formulation of Ash_{it}, β measures how monitors that are ever downstream may differ from nearby monitors in the same year, controlling for watershed monthly variation arising from seasonal factors like temperature. Variation in the the time-varying binary version of Ash_{it} may arise from plants shutting down, converting from coal to natural gas, or changing disposal practices. The time-varying binary version of Ash_{it} asks whether downstream monitors show differences in levels of water quality analytes compared to themselves in years when pollution sites are inactive. In this formulation, β is the average within-monitor difference in analyte level in years when upstream pollution sites are actively releasing compared to years when the upstream plants are not releasing water pollution. The final formulation of Ash_{it}, the tons released upstream in a year, varies due to plant closures and conversions and also from

[56] Watershed region refers to hydrologic unit (HU-6) geographies, which are watersheds roughly the size of an aggregation of several counties. See the USGS Watershed Boundary Dataset webpage for more information.

[57] Sample media include surface water, sediment, and hyporheic zone. Abnormal weather events include backwater, dambreak, drought, flood, hurricane, regulated flow, snowmelt, spill, spring breakup, and storm. Hydrologic conditions indicate whether the water level is low, high, or stable.

[58] In this procedure, monitor fixed effects are dropped, leaving only watershed-year and watershed-month fixed effects.

[59] Note that the TRI provides annual totals and not monthly pollutant loadings, so annual release quantities are merged to all months in any year.

[60] With multiple plants, the measure is calculated as: $Ash_{mt} = \Sigma_p 1[Downstream_m] * TonsReleased_{pt}$, where p represents a steam electricity generating coal power plant.

natural fluctuations in plant coal usage in a year. With this version of Ash_{it}, β estimates the relationship between each ton of coal ash released and the measured water property or concentration downstream.

Intuitively, Equation 1.1 captures how coal ash sites affect the properties of water downstream. It does so by comparing a specific location to itself in years when more or less is released upstream, controlling for local characteristics that may vary by month and year. The first formulation of Ash_{it} is not causal, although large and statistically significant differences across monitors in otherwise comparable regions may relate to the legacy of many decades of coal ash water pollution. Causal identification with the second and third formulations of Ash_{it} requires that no factors are correlated with the quantity of coal released and the property of water observed downstream, conditional on monitor, watershed-by-year, and watershed-by-month fixed effects.[61] Various concerns may arise with this estimation procedure. Previous studies demonstrate that standard statistical analyses are not ecologically relevant for physical and chemical properties of streams.[62] The same quantity of coal ash is likely to affect watersheds differently. Factors like total flow (and hence dilution), flow speed, temperature, agricultural activities, and tree coverage are all important determinants of how coal ash impacts a water system.[63] Moreover, these determining factors are likely endogenous to the quantity of coal ash released because regions with greater potential to absorb pollution may receive more of it. The monitor, watershed-by-year, and watershed-by-month fixed effects should allay some of these concerns. The prevalence of coal ash water pollution relative to other point-source pollutant categories also diminishes the likelihood that some other pollutant source might affect water quality to a similar extent.

---

[61] In Table B.12, I show that counties with coal ash pollution sites do not have statistically different quantities of water pollution or pollution impounded in landfills compared to other counties in the same state.

[62] Peterson et al. (2007).

[63] EPA (2015a).

4.2 Municipal Water Quality

Local geography, source water, system design, and homeowner characteristics influence municipal water quality.[64] [65] State regulatory monitoring tests report quantities across multiple facilities with different functions and monitoring requirements. State-level water quality regulations also play a role in observed water quality.[66] To determine the relationship between coal ash water pollution and municipal water quality, I address these factors with a municipal water system panel fixed effects specification. Consider the following regression:

$$y_{imst} = \beta Ash_{it} + X_{it}\gamma' + \eta_i + \eta_{st} + \eta_m + \varepsilon_{imst} \tag{1.2}$$

y_{imst} is the level of arsenic, conductivity, haloacetic acids, lead, pH, or trihalomethanes observed in municipal water system i, state-year st, and month m. Ash_{it} is the coal ash released into surface waters within 25 miles upstream of at least one of a municipal water system's intake locations in year t, where this value is replaced with zero if the Southern Environmental Law Center has not determined the water system to be sourcing from coal ash affected waters. η_i is a water system fixed effect, η_{st} is a state-year fixed effect, and η_m is a month fixed effect. I cluster all standard errors at the state and municipal water system level. X_{it} includes dummies for the facility type where the test occurred, system size dummies, system age, and a dummy variable equal to one if the analyte was not detected.[67] The facility type indicator controls for unobservable factors that differ across facilities within the same water system. η_{st} controls for any changes to state policies

[64] Gray and Shimshack (2011), Pieper et al. (2016).

[65] Water systems may use more than one source of water with differing underlying characteristics. For example, a system might have a groundwater well, a surface water intake, and also purchase water from a nearby system. Municipal water systems use different treatment techniques.

[66] Gray and Shimshack (2011).

[67] Facility types include distribution centers, transmission lines, treatment plants, source waters, wells, and homeowner tap-level tests. Time-varying system size dummies correspond to the five size categories used by the EPA to assign level of regulatory oversight to different systems. These categories are include very small systems (25-500 service population), small systems (501-3,300 service population), medium water systems (3,301-10,000 service population), large water systems (10,001-100,000 service population), and very large water systems (over 100,000 service population).

or secular pollution trends that may affect the levels of different compounds in a water system. Aside from a continuous measure of Ash_{it} representing the total tons released upstream, I test two alternative formulations of Ash_{it}. In the first, Ash_{it} is a simple binary indicating whether tons released upstream is positive, testing how water quality changes when a plant shuts down, converts, or changes pollution release practices. I also test a time-invariant version of Ash_{it} that is equal to one if the Southern Environmental Law Center determined the municipal water system is likely using water affected by coal ash pollution.[68] This formulation asks whether likely affected water systems are notably different from other water systems within the same watershed, conditional on state-year and monthly controls.

Intuitively, Equation 1.2 compares a municipal water system to itself in years with low or high upstream pollution releases. The coefficient β therefore estimates how an additional ton of coal ash water pollution released upstream in a year correlates with the level of a water quality indicator in a downstream affected water system. The identifying assumption of Equation 1.2 is that, conditional on water system characteristics, facility indicators, monthly fixed effects, and state-by-year regulatory changes, no factor is correlated both with the quantity of coal ash released upstream and the level of a specific pollutant in the municipal water system. This assumption may be violated if polluting firms near power plants systematically pollute similar compounds into surface waters in a way that is correlated with the quantity of coal ash effluent and the levels of an analyte in a municipal water system.

Next, I test the relationship between coal ash water pollution and the likelihood of a Safe Drinking Water Act (SDWA) violation. The Safe Drinking Water Inventory System tracks all municipal water system violations of the SDWA. I construct a panel of each water system in the inventory system for each year in which the system operated over 2000 to 2017, assigning violation counts by infraction type to each water system-year. For completeness, I employ both probit and linear probability models. Consider the following estimation procedures:

[68] In this formulation, I drop water-system fixed effects and add watershed fixed effects.

$$Pr(Vio_{it} = 1) = \Phi(\beta Ash_{it} + X_{it}\gamma' + \eta_i + \eta_t) \tag{1.3}$$

$$Vio_{ist} = \beta Ash_{it} + X_{it}\gamma' + \eta_i + \eta_t + \varepsilon_{it} \tag{1.4}$$

In Equation 1.3 and Equation 1.4, Vio_{it} equals one if water system i has a violation of the specific type in year t, and zero otherwise. I consider two types of violation outcome. In the first, I break up violations by type of infraction. In the second, I break up violations by the specific rule of the SDWA that was violated.[69] η_i is a water-system random effect in Equation 1.3 and a water-system fixed effect in Equation 1.4. η_t is a year dummy.[70] I cluster all standard errors at the municipal water system. X_{it} includes dummy indicators for five types of water system size, system type, owner type, school water system, surface-water sourcing water system, protected source-water, and water system age.[71] I test a time-varying binary and time-varying continuous formulation of Ash_{it}, as before. Equation 1.3 asks how being downstream from an active coal ash pollution site in a given year affects the probability of a water quality violation, or how each additional ton of upstream coal ash water pollution affects the probability of a water quality violation.

[69] Violations of the SDWA are laid out in the Safe Drinking Water Act by "rule" and "infraction." Rules include Arsenic, Consumer Confidence Rule, Filter Backwash, Disinfectant Byproduct, Groundwater, Lead and Copper, Miscellaneous, Nitrates, Public Notice, Radiation, Synthetic Organic Compounds, Total Coliform, Treatment Technique, and Volatile Organic Compound. Infractions against each rule include maximum contaminant level violation, monitoring violation, reporting violation, and treatment technique violation. Infraction types tend to vary by type of rule. For example, a consumer confidence rule is often related to reporting failures. A volatile organic compound violation may be related to monitoring lapses or, less commonly, maximum contaminant level violations. For each violation, an associated compound is listed. For example, a monitoring violation and a maximum contaminant level violation for the disinfectant byproduct rule may both list total trihalomethanes as the contaminant.

[70] I use the commands xtprobit , re and xtreg , fe in Stata.

[71] Federal types are community water system, non-community non-transient water system, and transient water system. Owner types are public and private, where public is the omitted category. School water systems are water systems that serve schools. Protected source-water indicates that a water systems source water is protected. I calculate age as the current year minus the date of first water system record in SDWIS. Note that many of these variables are dropped in Equation 1.4 because they are time-invariant.

4.3 Fetal Health

Unobservable factors are likely endogenous to household sorting across municipal water systems and hence water quality. Water quality violations, moreover, may present with simultaneous aversive behavior on the part of households.[72] I therefore model the relationship between coal ash water pollution and fetal health with a mother and zipcode panel fixed effects specification designed to control for time-invariant mother and neighborhood characteristics. Consider the following regression:

$$Health_{imtz} = \beta Ash_{it} + X'_{imt}\gamma + \eta_m + \eta_t + \eta_z + \varepsilon_{imtz} \tag{1.5}$$

$Health_{imtz}$ is a fetal health indicator for newborn i to mother m in year t and zipcode z. Health indicators include ounces at birth, low birthweight, preterm gestation, and presence of a congenital anomaly. η_m, η_t, and η_z are mother, year, and zipcode fixed effects. X_{imt} is a vector of time-varying birth and mother characteristics and county-level air quality measures.[73] X_{imt} also includes a dummy for whether the mother moved since the last observed pregnancy outcome.[74] I cluster all standard errors at the mother. Ash_{it} takes one of three forms. In the first, it is a time-invariant indicator applied to all water systems considered affected by coal ash according to the Southern Environmental Law Center. In the second, Ash_{it} is a binary variable indicating whether coal ash was released within 25 miles upstream of a water system's intake in year t. In the third, it is a continuous variable representing the tons of coal ash released within 25 miles upstream. Intuitively, Equation 1.5 estimates the difference in health outcomes across siblings where one sibling receives more potential exposure to coal ash water pollution. Such variation may arise from mother moves,

[72] Banzhaf and Walsh (2008), Bennear and Olmstead (2008), Zivin et al. (2011), Marcus (2019).

[73] Air pollution controls are mean, maximum, and maximum PM 2.5 squared in the county of residence across all months of gestation. Birth-specific controls include gender of the newborn and dummies for birth order. Mother-specific controls include age at time of birth, age squared, six dummy bins for number of clinic visits during gestation, and an indicator for tobacco use during gestation.

[74] For example, this variable equals one if the observed residence in period $t-1$ is different from the observed residence in t.

plant closures or plant conversions, and random variation in the quantity of water pollution in any year. In the time-invariant version of Ash_{it}, the identifying assumption is that mother's moves from or to coal ash affected regions are not associated with unobservable improvements in mother's well-being that may also affect fetal health conditional on controls for zipcode and the dummy indicator for having moved since the last pregnancy. In the second and third formulations of Ash_{it}, identification requires that factors correlated with plant closure or the quantity of coal ash released do not independently affect fetal health outcomes, conditional on controls for mother, zipcode, and year. A potential violation of this assumption would be if plant closures are associated with economic changes to the community that may affect mother health. Alternatively, a violation of the identifying assumption might occur if mothers systematically avert exposure to pollution in years when plants are active or when more pollution is released.

The primary source of variation in Equation 1.5 is mother moves. I therefore dis-aggregate the equation by mothers moving into and mothers moving out of coal ash-affected municipal water system service zones. Consider the following regression:

$$Health_{imtz} = \beta_1 \mathbf{1}[InMove_{it}] + \beta_2 \mathbf{1}[OutMove_{it}] + X'_{imt}\gamma + \eta_m + \eta_z + \eta_t + \varepsilon_{imtz} \tag{1.6}$$

$Health_{imt}$, X_{imt}, η_m, η_t, and η_z are as before. Rather than Ash_{it}, I now include two indicator variables capturing whether a newborn has been differentially exposed to an affected water service zone in comparison to its siblings. $\mathbf{1}[InMove_{it}]$ equals one if the listed residence of newborn i to mother m is within a coal ash affected water service zone while the listed residence for *previous* newborn j to mother m is *not* within an affected service zone. Conversely, $\mathbf{1}[OutMove_{it}]$ equals one if a newborn's listed residence is not within an affected municipal water service zone, while the listed residence of a previous newborn to the same mother was within an affected system's service zone. In cases where all children of the same mother are either exposed or not exposed, these indicators equal 0 for all newborns. In any case where multiple children are born after a transition to or away from an affected service zone, all subsequent children receive the same indicator.[75] Equation 1.6 includes an indicator for whether a mother moved since the last pregnancy to control for unobservable factors associated with moving that may also affect fetal health. Intuitively, Equation 1.6 asks whether observable fetal health differences may arise from moving into or moving out of coal ash water service zones.

4.4 Willingness-to-Pay for Avoiding Coal Ash Contamination

During a weather event in February of 2014, an ash pond along the Dan River in North Carolina burst its banks, releasing 25 million tons of coal ash into the nearby river. By September, the state legislature had responded with the Coal Ash Management Act, Senate Bill 729, in an effort to better manage coal combustion wastes. As part of the legislation, homes within 500 feet of a coal ash pond received mandatory home well water quality tests, where applicable. Many of these homes were found to have water considered unsafe to drink by the EPA.[76] Duke Energy subsequently provided these homes with bottled water for drinking and cooking. I test how this event, which led to information disclosure about well water quality and provision of bottled water by Duke Energy, affected home prices near the ash ponds. Consider the following equation:

$$y_{it} = \delta treat_i * post_t + \lambda post_t + \eta_i + \eta_t + \varepsilon_{it} \tag{1.7}$$

y_{it} is the sale price for home i in year t, where all prices are converted to 2014 dollars. Let $treat_i$ represent homes that are within a 1, 2.5, or 5 mile buffer region surrounding a coal ash pond. $post_t$ is a dummy equal to 1 if the sale occurred after 2014. $treat_i * post_t$ is the interaction of a dummy for the post period and an indicator for being within the circular buffer surrounding a coal ash pond. η_i is a fixed effect for either the home or the incorporated city of the home, and η_t is a set of year

[75] For example, a mother has one child in an affected region, and subsequently the mother has three children in an unaffected service zone. All three subsequent children receive an indicator of one for $\mathbf{1}[OutMove_{it}]$

[76] For more information, see this NC Department of Environmental Quality series of reports summarizing testing.

dummies.[77] I cluster standard errors at the county level in all analyses. The coefficient of interest in Equation 1.7 is δ, the average change in sale price of affected homes after 2014. The Dan River spill and the Coal Ash Management Act of 2014 made water quality concerns more salient by informing households of their well water quality, although households may have also adjusted risk perceptions due to the recent spill. Duke Energy began providing bottled water to affected residents at the same time as these other events. δ should therefore be interpreted as a change resulting from a variety of factors rather than one causal mechanism. Comparing sale prices to previous sale prices of the same home controls for time-invariant factors that may be unique to homes and neighborhoods near large power plants. Models using fixed effects at the city level require the identification assumption that homes nearer to coal plants would have similar sale price trends as other homes in the same city in the absence of the well water information disclosure, which is a stronger assumption.

5 Results

5.1 Surface Water Quality

Table B.4 shows the results of the surface water analysis for arsenic, chromium, conductivity, dissolved oxygen, lead, pH, selenium, and temperature. For each outcome, results are split into three columns depending on the Ash_{it} variable used in the estimation. The first column is a time-constant version of Ash_{it}, testing baseline differences in analyte between exposed and unexposed regions within the same watershed. Columns (2) and (3) display the monitor-specific fixed effects specification in Equation 1.1. These results regress a time-varying variable for coal ash releases on the relevant analyte, where column (2) is a simple binary if the monitor is exposed to positive releases in year t and column (3) is the annual tons released upstream within 25 miles. The coefficient on arsenic in column (1) means that monitors ever exposed to coal ash pollution have 0.0863

[77] I rule out using county fixed effects due to the substantial heterogeneity between homes near coal plants and other residences in the same county, both in average sale price and sale price trend. See Figure A.14 for a trend comparison.

mg/L greater concentration of arsenic than similar monitors in the same watershed-by-month and watershed-by-year cluster. For comparison, the standard maximum level of arsenic in municipal drinking water is 0.01 mg/L. In column (2), the coefficient of 0.0576 on arsenic suggests that downstream monitors have nearly six more mg/L of arsenic in years when upstream coal ash plants are releasing water pollution than in years when upstream plants are not releasing pollution. Finally, the coefficient on arsenic in column (3) suggests that each ton of coal ash released increases levels of arsenic in downstream monitors by roughly 0.002 mg/L. Scaling this by the average quantity of coal ash effluent released into surface waters in any given year (i.e., 10 tons), this point estimate suggests that an average coal ash release site emits enough surface water pollution in a year to make nearby waters exceed drinking water standards two times over. Similarly, baseline levels of the pollutants chromium, lead, and selenium are all elevated in downstream water quality monitors within 25 miles, although these results are only statistically significant for selenium and arsenic. For selenium, the drinking water standard is 0.05 mg/L, suggesting that a typical coal ash release site increases nearby concentrations of selenium by less than half the safe drinking water standard, although selenium is known to bioaccumulate in fish populations. Point estimates in columns (2) and (3) are noisier and, for chromium and lead, actually negative. Since it is unlikely that increased pollution lowers levels of pollutants in nearby surface waters, these perverse results may stem from measurement error or unobservable factors such as shifts in testing priorities after coal plants stop polluting.

In the second panel of Table B.4, column (1) demonstrates that surface waters downstream from coal ash sites have significantly deteriorated water quality indicators compared to nearby non-downstream bodies of water. Conductivity is nearly 1600 $\mu s/cm$ higher in these waters, which alone is nearly one third the average level observed in all non-downstream waters and slightly less than half of the average baseline difference observed between affected and unaffected water quality monitors.[78] Likewise, affected regions have lower baseline dissolved oxygen levels than comparable unaffected regions by roughly one tenth the mean level across all water quality monitors.

[78] A $\mu s/cm$ is a micro siemen per centimeter, a standard measurement of specific conductance.

Lower dissolved oxygen affects fish habitats and recreation value of water systems, although it also decreases the rate of pipe corrosion in municipal water systems that source from these waters. pH and temperature, meanwhile, are both significantly elevated in water systems affected by coal ash. pH tends to increase because of the many calcium and silica compounds present in coal ash; this effect is evidently only partially mediated by acid rain. Temperature, meanwhile, increases because power plants circulate nearby water in the electricity generation process. Strangely, the column (2) point estimate on conductivity suggests that closure of a coal plant is associated with an increase in conductivity of nearly 300 $\mu s/cm$. Indeed, the time-varying variables in column (2) and (3) often differ from the hypothesized relationships between pollution and water quality indices. These strange results, in combination with those for chromium and lead above, suggest that the time-varying pollution release variables may noisily capture true changes in surface water pollution.

Collectively, results in Table B.4 suggest that surface waters downstream from coal ash sites differ substantially from other nearby unaffected surface waters, although I find mixed evidence on the extent to which these differences are driven by contemporaneous pollution releases. One potential explanation is that waters in these regions are naturally different from waters in other regions within the same watershed. One alternative explanation is that, over many decades of coal ash pollution, these waters have developed significantly higher conductivity and pH levels that are not greatly affected by the contemporaneous amount of pollution released. Yet another possibility is a measurement error issue; measurement error of coal ash pollution may relate to poor self-reported estimates on the part of coal ash effluent managers, but it may also relate to substantial undocumented leakage and seepage from coal ash pollution sites. Management of ash wastes might also become less stringent after closure of a coal plant. Selection bias from changes to monitoring priorities over time may also account for some of these findings. I also display the results of a variety of inorganic compounds typically associated with coal ash in Table B.11. In general, these results support the main findings in Table B.4. In particular, I find evidence that coal ash water pollution increases levels of antimony, mercury, and thalliun.

5.2 Municipal Water Quality

Table B.5 displays the results of estimation procedure Equation 1.2. In this specification, I estimate the relationship between coal ash water pollution releases and the results of regulatory monitoring tests in nearby municipal water systems. I split the analytes into three categories and show results associated with three types of treatment indicators. The analyte categories are disinfectant byproducts, inorganic compounds, and properties. I now include two new analytes not present in the surface water analysis presented in Table B.4.[79] These water quality analytes are trihalomethanes and haloacetic acid.[80] The three treatment indicators correspond to those employed in Table B.4. The first column, labeled downstream, shows baseline differences between water systems that are believed to be sourcing from coal ash affected waters by the Southern Environmental Law Center and those that are not. Columns (2) and (3) test how monitoring test results within the same water system change over time in response to variations in the quantity of coal ash released upstream. All models include state-by-year and month fixed effects to control for time-varying state regulations and monthly fluctuations in water quality. Column (1), instead of a municipal water system fixed effect, includes a fixed effect for the watershed HUC-6 region of a municipal water system's intake location, which is assigned using the procedure described in Appendix Section 8.

Municipal systems that are affected by coal ash release sites, in comparison to other water systems within the same watershed and state-by-year combination that are not affected, tend to have lower conductivity, pH, and haloacetic acids.[81] Lead levels appear slightly elevated in affected water systems but not statistically significantly so. No other coefficient in column (1) is statistically

[79] I also drop chromium, selenium, dissolved oxygen, and temperature because these outcomes are either not tested frequently in municipal water systems (i.e., chromium, dissolved oxygen, and selenium) or not relevant to human health (i.e., temperature).

[80] I analyze the relationship between coal ash releases and the two most common and most-frequently tested disinfectant byproducts, haloacetic acids and total trihalomethanes. Although at least 500 disinfectant byproducts have been identified, these two compose at least 94% of all disinfectant byproduct formation (58% TTHM and 36% HAA5). Since disinfectant byproducts form during the water treatment process, I do not show any analysis of these analytes in surface waters. See DHHS for more information.

[81] Lower pH tends to create more haloacetic acids, while higher pH tends to form more trihalomethanes. Consistent with the lower average pH in affected water systems, it then follows that trihalomethanes may be elevated and haloacetic acids lowered in affected systems. The statistically lower haloacetic acid levels are therefore likely an artefact of baseline pH differences. See DHHS for more information.

significant. The results in column (1) are likely primarily driven by the unique characteristics of water systems sourcing from waters affected by coal ash. As is shown in Table B.2, these water systems are substantially larger and therefore subject to increased regulatory oversight. In addition, a water plant affected by coal ash pollution is necessarily using some quantity of surface waters. Surface waters contain fewer metals such as arsenic than groundwater; these differences also lead to altered conductivity and pH profiles.

The time-varying regressions of upstream coal ash pollution on water quality indicators provide mixed evidence of a contemporaneous link between coal ash water pollution and municipal water quality. All coefficients for disinfectant byproducts in columns (2) and (3) are negative and statistically indistinguishable from zero. The point estimate in column (2) for arsenic suggests that a plant closing upstream is associated with a 0.0084 mg/L improvement in arsenic levels. Compared to the water quality standard for arsenic of 0.01 mg/L, this improvement is quite dramatic. However, it is not statistically significant at the 10% level of confidence. For lead, I find that each ton of upstream coal ash water pollution increases downstream municipal lead levels by 0.0035 mg/L. This is roughly one fifth the maximum contaminant level for lead, 0.015 mg/L. Since the average quantity of coal ash released upstream in any given year for municipal water systems is 4 tons, this is a sizeable increase in municipal lead levels.[82] Next, I find that conductivity tends to be higher in these water systems in years when upstream coal plants are active. In years when no pollution was released, conductivity in downstream water systems was 45 $\mu s/cm$ lower. Each ton of coal ash released, meanwhile, is associated with a 3 $\mu s/cm$ increase in conductivity in downstream municipal water systems. Interestingly, neither of these effects would close the gap in conductivity between coal ash affected and unaffected water systems.[83] This suggests that, although affected water systems have less-corrosive water on average, coal ash pollution may nevertheless lead to infra-marginal changes in water quality that affect pipe corrosion and tap lead levels without nec-

[82] Note that the average upstream releases for affected water systems differs from the average effluent release quantity across all TRI sites because municipal water systems are not always placed near pollution sites. Moreover, water system intakes are unlikely to be placed near the heaviest-polluting sites.

[83] Unaffected systems average 299 $\mu s/cm$ and affected systems 183 $\mu s/cm$. Recall that this difference is driven largely by use of groundwater sources by unaffected systems, and groundwater has higher conductivity.

essarily causing increased regulatory notice. In the last row of Table B.5, estimates for the effect of coal ash pollution on pH are of opposite sign. This discrepancy may relate to the influence of acid rain on water pH profiles.

In the next municipal water quality analysis, I test how changes in coal ash water pollution affect the likelihood of water quality violations of the Safe Drinking Water Act. These results are displayed in Table B.6 by type of infraction and type of rule.[84] I display all infraction types and rule types that are potentially related to coal ash water pollution.[85] In column (2), I find evidence that water systems experience fewer health-based violations, maximum contaminant level violations, reporting violations, arsenic violations, and inorganic compound violations in years when upstream plants are not polluting surface waters. As shown in row (1), these differences exist despite affected water systems having statistically similar violation rates in years when relatively more pollution is released into upstream surface waters. I find evidence that each ton of coal ash released upstream increases the likelihood of violations for disinfectant byproducts and inorganic compounds. In the case of inorganic compounds, I find that a one ton increase in upstream pollution releases increases the likelihood of a violation by 0.15 percentage points, which is a massive increase compared to the baseline annual violation rate for this rule of 0.24%. Conversely, an additional ton increases the likelihood of a disinfectant byproduct violation by only 0.1 percentage points, which is a smaller fraction of the annual violation rate for that rule, 1.34%. In all specifications, volatile organic chemicals appear positively associated with coal pollution releases, although in no specification are these relationships statistically significant. Puzzlingly, I find a negative and statistically significant relationship between plant cessation of water pollution and lead and copper violations. Results are generally consistent across OLS and probit models.

[84]Note that infraction types and rules are not mutually exclusive; infraction types are specific ways in which a water system might break a rule.

[85]The category "inorganic compounds" includes many potentially coal-associated compounds.

5.3 Fetal Health

In Table B.7, I present the results of Equation 1.5 across four measures of fetal health: birthweight in ounces, an indicator for low birthweight, an indicator for preterm gestation, and an indicator for the presence of any congenital anomaly. I show these results in three panels, each corresponding to a different formulation of Ash_{it} as laid out in Equation 1.5.

In Panel A, all point estimates are identified off mothers moving into or out of geographies served by municipal water systems using coal ash affected source waters. The coefficient in column (1) suggests that a newborn potentially exposed to coal ash pollution, in comparison to an unexposed sibling, is 1.2 ounces lighter. Such newborns, in comparison to their unaffected siblings, are 1.7 percentage points more likely to have low birthweight. They are also 1.3 percentage points more likely to be preterm. These newborns also appear slightly more likely to have a congenital anomaly, although this difference is not statistically significant. These differences in fetal health are large in magnitude relative to the baseline fetal health means across the state. They're also large relative to the effect of differential fine particulate matter exposure *in utero*. For example, the average difference in particulate matter exposure for mothers ever exposed to a coal ash affected municipal water service zone is 0.5 $\mu g/m^3$. In combination with the point estimate in on air pollution exposure in row (2), this discrepancy suggests that, from air pollution exposure alone, these mothers would be expected to have newborns roughly 0.55 percentage points more likely to be preterm. The same estimate associated with potential water pollution exposure is over twice as large. Meanwhile, mothers with less education, who are expected to be less able to avert water pollution exposure using water filters and other pollution aversion strategies, are more affected across all fetal health indicators except congenital anomalies. All estimates are also conditional on zipcode fixed effects and an indicator for moving since the previous pregnancy, which control for potential changes in life circumstance that may be associated with mother moves.

In Panels B, column (1), I show that cessation of upstream water pollution practices is associated with a decrease in infant weight of 0.41 ounces. This result is highly significant. Moreover, after cessation of upstream coal ash water pollution, newborns appear slightly less likely to have low birthweight or be preterm. Panel C provides qualitatively similar estimates of different magnitude; an additional ton of coal ash released upstream is associated with an improvement in fetal birthweight of 0.014 ounces. These results are puzzling; it is certainly unlikely that more pollution would improve fetal health. Therefore, it seems likely that unobservable factors associated with coal plant operation and pollution also affect fetal health. For example, closure of a coal plant might change local economic conditions in a way that affects fetal health. Perhaps more likely, aversive behavior on the part of mothers might attenuate and even reverse the potentially negative impacts of coal ash water pollution. Mothers might be more likely to drink bottled water or purchase home filtration devices in years with more pollution or active nearby pollution site, even though the legacy of coal pollution may linger after cessation of active coal ash effluent releases.

Because the results in Panel A are driven exclusively by mother moves, I also disaggregate these effects by moves into or out of coal ash affected municipal water system service zones in Table B.8. Mothers moving into affected service zones have newborns that are, in comparison to previous newborns, 1.8 ounces lighter. These affected newborns are also 2.8 percentage points more likely to have low birthweight, and they are 2.1 percentage points more likely to be preterm. Mothers moving out of coal ash affected regions, meanwhile, see their newborns increase in birthweight by 0.58 ounces, although this difference is not statistically significant.[86] Similarly, out-movers see improvements in the likelihood of having low birthweight of one percentage point. Out-movers also appear to dramatically lower the likelihood of a congenital anomaly; this improvement should be prefaced with concerns about the congenital anomaly indicator discussed in section 3.4. For nearly all outcomes, point estimates suggest that moving into an affected service zone worsens fetal health, while moving out of one improves it. Since the association between tobacco use during pregnancy and low birthweight is roughly 4 percentage points, the increase in incidence of low birthweight in Table B.8 of 2.8 percentage points is a dramatic change.[87]

[86] Among other potential explanations for the divergence in effects across in-movers and out-movers, it is possible that mothers moving out of coal ash affected areas carry with them the cumulative effects of previous pollution exposure.

[87] Zheng et al. (2016)

5.4 Home Sale Prices

Table B.9 shows how sale prices of homes near coal plants changed in North Carolina after 2014, the year of a large coal ash spill and the state's Coal Ash Management Act. The first three columns show the results of estimation procedures with city fixed effects; columns (4) to (6) show the results with home fixed effects. Homes within 1 to 5 miles of coal ash ponds experienced sale price decreases of 5% to 14% after 2014, depending on the distance cutoff and comparison group. Models with home fixed effects have smaller point estimates across all distance bandwidths, suggesting that within-city comparisons may confound differential trends of the comparison homes with the policy. All models, however, suggest large, negative, and significant sale price changes. Sale price declines of homes within one mile of a coal ash pond are between 12% and 14% depending on the type of fixed effect employed, which is a substantial decline in homeowner wealth. Homes closest to ash ponds experienced the largest changes in sale price, with the effect size decreasing monotonically with distance from the coal ash ponds. The price changes may relate to increased salience of coal pollution, the dis-amenity value of recently-discovered unsafe well water, or changing secular preferences for pollution after the Dan River spill.

5.5 Cost Analysis

I perform back-of-the-envelope calculations of the external cost of coal ash water pollution with respect to two outcomes: low-birthweight newborns and changes in home sale prices. In Table B.7, the coefficient of 0.017 implies that mothers served by municipal water systems affected by coal ash are 1.7 percentage points more likely to have a child of low birthweight. This implies roughly 700 additional newborns of low birthweight.[88] 700 low-birthweight newborns is approximately 0.5% of the total of low-birthweight newborns in North Carolina from 2005-2017. These low-birthweight newborns likely led to $10.7m in additional hospitalization fees and $2.8m in K-

[88] 900,000 of 1.5m newborns in the sample are served by municipal water systems, and 1 in 22 are served by municipal water systems affected by coal ash. 0.017*900,000*$\frac{1}{22}$ is 695.45.

12 educational expenses for local communities.[89] These costs do not account for many additional expenses associated with low-birthweight newborns, such as later-life health complications or increased social services excluding special education. As for real estate, Table B.9 presents likely total changes in home sale value associated with the revelation of non-potable drinking wells in homes surrounding ash ponds. These estimations multiply the per-home change in sale price by the number of homes affected in each distance cutoff. Results suggest likely changes in home values between $20 million and $450 million, depending on the model and distance cutoff.

6 Policy Relevance

The Environmental Protection Agency recently promulgated two rules with respect to the management of coal ash waste. The first, known as the Effluent Limitation Guidelines, stipulates that certain types of coal ash waste are not to be released into surface waters and that ash pond effluent streams must not exceed limitations on the concentration of specific compounds.[90] The second rule modifies subtitle D of the Resource Conservation and Recovery Act, which allows the EPA to regulate pollutants from cradle to grave. Known as the Disposal of Coal Combustion Residuals from Electric Utilities Rule, it establishes requirements for surface impoundments receiving coal ash wastes; among other stipulations, the rule mandates structural integrity tests, groundwater monitoring, run-off controls, and record keeping requirements. The rule also creates new guidelines with respect to the closure of inactive coal ash impoundments. These rules reflect current understanding of best practices for the management of coal ash waste. The Effluent Limitation Guidelines, in particular, are estimated to decrease the quantity of coal ash that may affect surface

[89] These numbers generated assuming each low-birthweight newborn costs an extra $15,000 and that each low-birthweight newborn is twice as likely to qualify for special education, with costs of roughly $44,000 per student. I assume baseline likelihood of special education service provision is 10%. Cost estimates from Petrou (2003) and Russell et al. (2007). Note these estimates are based on associational evidence.

[90] Managed waste types include many relatively new forms of coal ash waste generated in larger quantities due to technical changes in the way that coal ash and coal-related air pollution are managed. For example, installation of scrubber technology creates flue gas desulfurization waste. See the Technical Development Document (EPA, 2015c) for more information.

waters by at least 95%.[91] However, the estimated benefit-cost ratios for the Effluent Limitation Guidelines are not always greater than one.[92] This study provides novel evidence that additional public health benefits from improved municipal drinking water quality are probable and likely economically meaningful.

Other policy levers may also ameliorate the potential influence of coal ash pollution on nearby surface waters, municipal water quality, and exposed populations. Remediating an older ash pond by treating the water, excavating the ash, and moving the ash to a new location is one such option; cleaning an ash pond has immediate effects on groundwater, improving arsenic levels by as much as 90 percent.[93] Such ash pond remediation, however, can be very expensive.[94] Increased recycling of coal ash into fertilizers and concrete, already commonplace, could also be expanded to reduce the environmental footprint of this waste.[95] For concerns related to the burden of payment for cleanup, local legislative acts have also been passed that prevent recuperation of costs from illegal coal ash discharges.[96]

7 Conclusion

I find evidence that coal ash surface water pollution affects nearby surface water quality. Discharges of coal ash are associated with increased conductivity and pH in downstream surface waters and municipal waters sourced from the same locations. These changes are driven in part by contemporaneous pollution releases, as heavy metal compounds found in coal ash are also found in higher concentrations in affected waters in years when more pollution is released. Differences in fetal health across siblings provide evidence that this pollution matters for human health, especially for mothers with less education who may be less able to avert pollution. Revelation of groundwater

[91] EPA (2015a).

[92] EPA (2015b).

[93] Fretwell (2016).

[94] In North Carolina alone, the cleanup is expected to cost in excess of $10 billion. https://www.utilitydive.com/news/duke-north-carolina-coal-ash-pond-excavation-order-to-cost-4-5b/551788/

[95] Yao et al. (2015).

[96] https://www.ncleg.net/Sessions/2013/Bills/Senate/PDF/S729v6.pdf

contamination decreased home sale prices in regions near coal plants in North Carolina across all models and specifications. Back-of-the-envelope calculations suggest substantial external costs of this form of pollution, which are likely understated.

8 Appendix

8.1 Assigning Downstream Status to Monitors and Water Systems

The National Hydrography Dataset Plus (NHD) is a GIS database of every water network in the United States. It features "edges," or river system segments and polygons identified by their COMID identifier, and "nodes," or midpoints of river system segments or polygons. I use the STARS package, an ArcGIS add-on, to assign coal ash release sites to river system edges in the NHD using the snap tool.[97] I then trace out downstream segments using the downstream tool, which creates polygons for the downstream regions from each coal ash release site. I then calculate distance downstream from each coal ash plant for each river edge, allowing sites with multiple upstream coal ash plants to have at least two unique observations. All monitoring locations in the Water Quality Portal are then joined by nearest spatial location to edges in the NHD. This allows merging river edge information on coal ash releases to water monitoring sites located on those edges. I can then calculate the total quantity of upstream coal ash released across different distance cutoffs, or weight the quantity released by the distance to each plant.

8.2 Assigning Municipal Water System Location

Performing an analysis of the relationship between water pollution and municipal water quality requires relatively accurate placement of wells and intakes. Due to security reasons, the location of these wells or intakes is typically not published online or accessible.[98] Moreover, municipal water systems often have wells or surface water intakes that are many miles away from their service zone, and larger systems typically have many intake locations. To assign municipal water systems to water source locations, I rely on three datasets and multiple linking procedures. First, I secure North Carolina's public water ground- and surface-water supply shapefile.[99] To this, I then add

[97] Peterson and Hoef (2014)

[98] A notable exception is North Carolina, which makes available all municipal water system intake locations as a geographic shapefile through its NC Onemap service. However, conversations with state water system planners suggest that these locations are published with some imprecision for security reasons.

[99] See here to download or see more information.

the Southern Environmental Law Center's public water system intake geodatabase, which shows surface-water intake locations for Alabama, Georgia, Tennessee, and Virginia. In some cases, these locations are approximated. Both well and surface water intake locations are included in the SELC database for North and South Carolina. Intake locations accessible online and the SELC geodatabase do not include many intake locations over the remaining states and even some within North and South Carolina. I supplement these data by approximating the remaining intake locations using the Safe Drinking Water Inventory System (SDWIS). SDWIS provides water system addresses, but these addresses are inaccurate. They represent the location of the water system managing office or long-distance owner.[100] I therefore approximate intake location based on service zone city or zipcode, county, and state. I then spatially join these locations to the nearest "downstream" polygons of river segments, excluding any link with a distance greater than 75 kilometers. The assumption is that any link greater that 75 kilometers away is very likely not using, purchasing, or otherwise influenced by the downstream water segment. I only use these approximated locations in instances where the intake or well location is not already known. These linking procedures allow me to approximate upstream pollution releases for any water system, although I only assign these upstream variables in cases where the SELC determined a water system to be using coal-ash affected waters. Water systems that may appear downstream on a map but that are known to source their water from a protected source are therefore not considered treated by coal ash water pollution in this analysis.

8.3 Assigning Air and Water Quality to Birth Residences

The North Carolina State Center for Health Statistics provided residential address information for all births in the state. These addresses included county information, which is used to assign air quality information at the county-month level to each birth. Since a birth is potentially affected by air quality across its entire gestational period, I assign mean and maximum PM 2.5 to each

[100]For example, some water system addresses were in California and New York State, while others were located in larger cities within the same state but hundreds of miles away.

birthday-county-gestation length combination. The mean fine particulate matter control is the mean level observed in the county over the entire gestational period, while the maximum value is the maximum county-month value over the entire gestational period. Averaging over the entire gestational period allows children with the same birthday and county of residence to potentially have different air quality controls if their gestational length differs. For example, a birth with gestation length of nine months receives a particulate matter control of the average of each of the nine months prior to birth, while a birth in the same county in the same month with gestational length of eight months will have a mean particulate matter control constructed over a different time period. Likewise, the maximum particulate matter control, the highest monthly average PM 2.5 observed during the entire gestational period, could differ across births within the same county and month if gestational length differs.

Assigning gestational periods to water quality information first requires linking residences to municipal water service zones. I therefore geo-code a statewide property parcel database to geographic shapefiles of all municipal water service zones. After linking these addresses to service zones, I string match the addresses listed in the birth certificates database to the addresses in the state parcel database using the Stata program matchit.[101] Next, I list out all North Carolina cities associated with coal ash sourcing water systems according to the Southern Environmental Law Center, and I merge any unmatched births to these city-water system combinations where applicable. After these steps, roughly 700,000 of 1.6 million birth residences are matched. Finally, I create a variable for the mode municipal water system by zipcode, and I replace any missing water system links with the mode water system for that zipcode. Because this imputation procedure is likely imperfect, I flag these imputed water system links and control for the imputation in all birth regressions. After all merges are complete, 1.1 million birth residences are linked to municipal water systems. Since roughly two thirds of individuals in North Carolina use municipal water and the remainder use home wells, the linkage procedure assigns roughly the correct proportion of addresses to municipal water systems.

[101] Raffo (2015).

Chapter 2

School Bus Emissions, Student Health and Academic Performance

1 Introduction

Nearly 25 million children ride over 500,000 buses to school in the United States each day. The predominantly diesel bus fleet contributes to air pollution exposure that may adversely affect children's health and academic performance. Because of this, school bus retrofit programs have been enacted across the country, making use of up to $200 million in federal grants per year to local districts to replace or retrofit engines. We use information on 2,656 of these school bus retrofits in Georgia, affecting approximately 150,000 students, to estimate effects on student health and academic achievement.

Diesel retrofits are an immediate and relatively inexpensive way to dramatically reduce diesel emissions.[1] A large literature has estimated the effect of diesel engine emissions on ambient air quality, in particular on nitrogen oxide and particulate matter.[2] A separate literature examines the effect of exposure to air pollution on children's academic achievement and health.[3] Yet, little is

[1] Barone et al. (2010), Tate et al. (2017).
[2] EPA (2009).
[3] Lavy et al. (2014), Currie and Neidell (2005), Marcotte (2017).

known about the direct effect of diesel emission reductions on children's academic achievement or health. The only studies to investigate school bus retrofits on health outcomes are Beatty and Shimshack (2011), which finds that bus retrofits in Washington state lead to significant reductions in asthma and pneumonia doctor visits, and Adar et al. (2015), which finds that retrofits in Washington state reduce pollution and pulmonary inflammation and increase lung growth. No study we know of examines the effect of reduced exposure to school bus emissions on academic performance.

To address the causal link between diesel retrofits, student health and academic achievement, we exploit variation in the timing and location of over 2,600 school bus retrofits across Georgia between 2007 and 2015. During our sample period, 15 percent of Georgia's 180 school districts retrofitted a share of their fleet. Our measure of exposure at the district level is based on the proportion of the bus fleet retrofitted in a given district. We further refine this with the proportion of students who are bus riders and the average amount of time students spend on the bus. We match retrofitting data to two types of district-level outcome measures: student health and scholastic outcomes. For the former, we observe a state-mandated fitness evaluation known as FitnessGram.[4] These health data include an established measure of cardiovascular health (aerobic capacity), which allows us to estimate effects on respiratory health, and BMI, which we take as a potential placebo against general health trends, though we discuss why BMI might also be affected by improved respiratory health. For scholastic outcomes we observe English and math end-of-grade test scores in addition to attendance.

We find positive and non-trivial effects of bus retrofits on student health. Retrofitting an entire fleet leads to a 4 percent increase in the average aerobic capacity of students, or roughly 1.8 units of VO_2 max, in our most conservative estimate. This effect is slightly larger when we weight treatment by the share of students in a district who ride the bus. In this case, retrofitting 100 percent of buses in a district where everyone rides the bus would yield a 5 percent improvement

[4]The FitnessGram© tests have been used for decades to assess student health, and a large literature demonstrates the scientific validity of the tests employed. The FitnessGram manual (https://www.cooperinstitute.org/vault/2440/web/files/662.pdf) provides details.

in aerobic capacity. We find no relationship between retrofits and our placebo, BMI. We show that effects on aerobic capacity are strongest for elementary school students.

We also find evidence that these retrofits affected student achievement. Retrofitting 10 percent of a district's fleet increases English test scores by 0.009 standard deviations, so retrofitting an entire district's fleet would increase test scores by nearly one-tenth of a standard deviation. Weighting by the share of students who ride the bus, we find that districts experience a 0.14 standard deviation increase from retrofitting an entire fleet when all students ride the bus. Estimated effects on math scores are also positive, but are smaller and noisier than those for English and often cannot be distinguished from zero. We find little evidence that attendance was significantly affected, though initial attendance rates were very high.

Our results suggest that retrofits are a cost-effective lever to improve both student health and achievement. A back-of-the-envelope analysis suggests that for each effect, benefits were far in excess of costs. The average retrofit required only $8,110 in our sample, suggesting diesel engine retrofits can be at least three times more cost-effective than class-size reductions for achieving a given test score improvement.

2 Background

School bus diesel emissions are a public health concern because school buses are ubiquitous, concentrated in residential areas, and dirtier than most vehicles. Monahan (2006) finds that California school buses were nearly twice as polluting as the average tractor-trailer. This is primarily due to the age of the bus fleet; a 30-year-old school bus can produce two or three times as much on-board pollution as a 3-year-old bus.[5] School buses are also exceptionally dirty because diesel emissions are more polluting than gasoline emissions, contributing to a third of nitrogen oxide emissions and a quarter of particulate matter emissions despite being a smaller fraction of the automobile

[5] Harder (2005).

fleet.[6] [7] School buses contribute to pollution exposure both for individuals spending more time near bus stops and along bus routes, but they are highest for passengers of the vehicle.[8] In fact, Zuurbier et al. (2010) find that riders of diesel buses had twice as much exposure to air pollution as carpoolers.

2.1 Emissions and Health

Exposure to air pollution worsens infant and childhood health. Diesel emissions contain smoke-related particulate matter, nitrogen dioxide, gaseous aldehydes, carbon monoxide, and toxic poly-cyclic hydrocarbons. The latter are potent carcinogenic compounds that are more stable when they diffuse into airborne water vapor, allowing them to reach deep into the lungs when inhaled.[9] For this reason, diesel exhaust may cause immediate short-term adverse pulmonary effects by decreasing the membrane potential of epithelial cells in the lungs.[10] There are also longer-term effects of diesel exhaust exposure; one cohort study of urban bus drivers in Denmark finds that just three months of bus driving is associated with an increased risk of six types of organ-based cancers and all malignant tumors.[11] Young individuals are especially vulnerable to this form of pollution. Worse air quality is linked to child lung function growth disparities of 3 to 5 percent, or four times the effect of second-hand cigarette smoke, in more-polluted areas, while exposure to in-traffic air pollution is associated with lower lung capacity, lower forced expiratory flow, and asthma development.[12] Two recent studies exploit variation in bus pollution at the census block level in New York City. The first (Ngo, 2015) finds that increasing emission standards over time reduced emergency department visits for respiratory diseases among residents living within a few hundred feet of a bus route. A second (Ngo, 2017) exploits variation in bus age, and thereby pollution levels, finding

[6] EPA (2009).
[7] In our sample, 99% of school buses are diesel-powered.
[8] Xu et al. (2016), Marshall and Behrentz (2005).
[9] Commins et al. (1957), Waller et al. (1985), Muzyka et al. (1998).
[10] Stevens et al. (2010).
[11] Soll-Johanning et al. (1998).
[12] Beatty and Shimshack (2014), Gauderman et al. (2005), Gendron-Carrier et al. (2018), Clougherty and Kubzansky (2008).

that children born to mothers who lived close to bus routes with older (dirtier) buses see modest reductions in infant birth weight and gestational age compared with those living near routes with newer, cleaner, buses.

2.2 Emissions and Academic Performance

Past work has identified three mechanisms through which air pollution may impact test scores: attendance changes due to pollution-related illness, short-term disruptions in attention and cognitive performance, and long-term negative influence of pollution exposure on brain development. Currie et al. (2009) demonstrate that higher pollution levels over six-week periods are associated with more student absences, which may indirectly impact student learning. Marcotte (2017) finds that daily pollen and particulate matter pollution levels affect students' math and reading test scores. Ultrafine particles in air pollution, particularly in diesel emissions, deposit in the prefrontal cortical and subcortical regions of the brain via the olfactory bulb, leading to heightened inflammatory response, white matter lesions, and behavioral and cognitive impairment.[13] Such cognitive impairment is observable in standardized test scores, and the negative effects stem from both contemporaneous and long-term exposure.[14]

2.3 Emission Reduction Programs

The well-known dangers of pollution from school bus diesel emissions led the United States Congress to spend $200 million per year from 2007-2012 to retrofit buses under the Diesel Emissions Reductions Act. Separately, the Clean School Bus Grant Program spent $110 million in 2005 and 2006. These grants pay for any one of four types of engine retrofits in our sample: diesel particulate filter, diesel oxidation catalyst (DOC), flow-through filter, or a closed crankcase filter (also called a closed crankcase ventilation system or CCV). Since the average diesel particulate filter costs between $5,000 and $10,000, engine retrofits have the potential to be a cost-effective means

[13] Freire et al. (2010), Guxens and Sunyer (2012), Calderón-Garcidueñas et al. (2012), Sunyer et al. (2015).
[14] Ebenstein et al. (2016), Chen et al. (2017), Ham et al. (2014), Marcotte (2017).

of reducing ambient air pollution and the health concerns associated with them.

The most common type of retrofit, a diesel particulate filter, can decrease overall emissions of particulate matter (PM) between 60 and 90%.[15] The effect of these filters on PM levels inside the bus cabin is more modest at between 15-26%.[16] Emissions reductions of heavy metals from a diesel particulate filter are more substantial, in the range of 85-95%.[17] Emissions of other harmful compounds, such as total hydrocarbons and carbon monoxide, can be reduced to background pollution levels.[18] Finally, reductions of nitrogen oxide emissions can be significant; Tate et al. (2017) find that retrofitting the bus fleet in York, UK, would reduce city-wide levels of nitrogen oxides by 6-7%. These benefits appear to be fairly persistent with good engine maintenance and the use of low-sulfur fuels. Another study finds that the reductions in PM of 95% by mass remained after four years of road exposure.[19] Taken together, the existing scientific evidence suggests that retrofits dramatically reduce the exposure of students to potentially harmful compounds.

Our work builds most directly on Beatty and Shimshack (2011), who examine roughly 4,000 school bus retrofits in Washington state between 1996 and 2006. They match retrofit data and hospital admissions at the district-month level. The authors find that districts with retrofits see significant and sizable reductions in asthma and pneumonia-related visits for both children and adults, with estimated benefits of nearly 7 to 16 times the cost of retrofit investments. In a related article that focuses on direct measures of exposure to pollution, Adar et al. (2015) measure pollution and health of 275 elementary school bus riders in Seattle and Tacoma, Washington, during a retrofit program from 2005-2009. The authors separately estimate the effect of four different emissions reduction programs (DOCs, CCVs, and fuel switching to ultra-low-sulfur diesel or biofuels) on pollution exposure, health measures, and school absenteeism. They find significant effects of DOCs, CCVs, and ultra-low-sulfur diesel use on on-board particulate levels. They find health ben-

[15] Biswas et al. (2009), EPA (2009).

[16] Hammond et al. (2007).

[17] Hu et al. (2009).

[18] Jiang et al. (2018). Note that Zhang and Zhu (2011) find that retrofits significantly decrease tailpipe emissions but have no significant effect on on-bus ambient air quality, while Li et al. (2015) show that tailpipe emissions do in fact enter the cabin. Borak and Sirianni (2007) conduct a meta-analysis and conclude that control technologies like retrofits can in fact eliminate "self-pollution" from diesel exhaust into bus cabins.

[19] Barone et al. (2010).

efits (increased lung functioning measures) from DOCs and CCVs only for students with persistent asthma.

We build on this prior literature in several ways. First, we have different measures of student health: aerobic capacity and BMI from FitnessGram tests. Beatty and Shimshack (2011) use hospital visits, and Adar et al. (2015) use measures of lung functioning.[20] Our health outcome measures are likely to better capture the effect of diesel emissions on student health because VO_2 max conveys general cardiovascular health rather than lung function, therefore representing the observable consequence of lower lung functioning. Our outcome also captures the health of all students instead of merely those visiting a clinic for acute lung conditions, thereby capturing the effect on the average student instead of only those likely to visit a clinic. Second, we provide potential placebo measures using a non-respiratory health outcome, BMI. Third, ours is the first study we know of to examine the effect of retrofits on academic performance, allowing us to tie together two largely separate literatures on health and academic performance.[21]

2.4 Retrofits in Georgia

The Georgia retrofit program started as the Adopt-a-School Bus program in 2003, a collaboration between the state Environmental Protection Division, school districts, and businesses to improve the well-being of students. The goals of the project were to implement any of four emission reduction retrofit devices, reduce bus idling, and increase use of ultra-low sulfur diesel.[22] The project has since been funded by a wide variety of sources and grants. The EPA Clean School Bus grant program provided three separate grants in 2004, 2005, and 2006. The Diesel Emissions Reduction Act (DERA) was passed by Congress in 2005 as part of the Energy Policy Act and is administered

[20] Adar et al. (2015) use forced expiratory volume in one second (FEV1) and forced vital capacity (FVC) as measures of lung functioning. These measures are useful figures for diagnosing lung diseases such as COPD or emphysema, but they do not measure cardiorespiratory fitness *per se*. The FitnessGram aerobic capacity test we employ is designed to capture VO_2 max, the maximal oxygen uptake at peak performance. Other studies, including Ross et al. (2016), use VO_2 max as a broader indicator of health.

[21] By contrast, other studies, including Marcotte (2017), estimate the effect of pollution exposure on academic outcomes but do not conduct a program evaluation.

[22] Idling reductions were a statewide effort.

by the EPA. Under DERA, the EPA sponsored two retrofit grant cycles in 2009 and 2014 that collectively paid for 182 school bus retrofits. The US Department of Transportation sponsored the program under its Congestion Mitigation and Air Quality Improvement (CMAQ) Program, which contributed $11.2M to retrofit 1,890 buses. The staggered funding and implementation lags allow us not only to compare retrofitting and non-retrofitting districts, but also to exploit the timing of retrofits among retrofitting districts to secure causal identification.

Over the relevant sample period from 2007-2017, 2,656 buses were retrofitted with at least one type of modification. 1,160 of these bus retrofits involved a diesel particulate filter, 1,394 added a diesel oxidation catalyst, 58 installed a flow-through filter, 244 added a closed crankcase filter, and 188 buses were replaced early. We do not observe any information on the use of ultra-low-sulfur diesel (ULSD) fuel, but we know from communication with the Environmental Protection Division that retrofit grants stipulated the use of ULSD fuel to preserve the new engine parts. Moreover, EPA diesel fuel standards required the use of ULSD on all vehicles starting in 2010.

3 Data

Our data come from four sources, providing information on health, achievement, retrofits, and the Georgia bus fleet in general. Since we observe school bus retrofits at the district level, we aggregate data to that unit of analysis. We describe each data source, advantages, and limitations in turn below.

3.1 Health

Our first data source contains health information from the Cooper Institute's FitnessGram examination. The FitnessGram examination is a series of mandatory tests administered annually to all Georgia public school students who are in a physical education class. Many other states use FitnessGram as well, and the results of the FitnessGram tests are used widely in studies on student health.[23] According to the Georgia Department of Education's 2016 Fitness Assessment Program Report, 1.1 million students in Georgia (74%) participate in the examination. Since physical education requirements differ by age, the participation rate for elementary school students is 94%, while for middle school and high school students the rate is 71% and 49% respectively. Since our study covers several years, most students should be included at some time in the sample window.

Several tests are involved in a FitnessGram examination, including tests of aerobic capacity, body mass index, curl-ups, push-ups, and sit and reach (a measure of flexibility).[24] We limit our analysis to just the test of aerobic capacity and BMI. Aerobic capacity is a measure of cardiovascular fitness likely to be affected by exposure to diesel pollution, and BMI is a potential placebo.[25]

Aerobic capacity is the maximum rate at which oxygen can be taken up and utilized by the body during exercise. It is measured by FitnessGram through an exercise called the PACER (Progressive Aerobic Cardiovascular Endurance Run) test, also called a multi-stage fitness test, a "beep test", or a shuttle run.[26] Physical education instructors administer the test and record results according to instructions provided by the Cooper Institute. The school-level average VO_2 max, as computed from either the student-level number of laps completed on the PACER test or the timed performance on a one-mile run, is our observed outcome measure.[27] The FitnessGram assessment

[23]Castelli et al. (2007), Welk et al. (2010), Edwards et al. (2011), Fahlman et al. (2006), Murray et al. (2012), Anderson et al. (2018).

[24]Records of these assessments are kept by the Georgia Department of Education Physical Fitness Division, which annually reports school-level results separately for male and female students. For each school-gender-test combination, measures include the total number of attempts, the average performance, and the percentage of students attaining "healthy fitness zone" (HFZ) status. Depending on whether the aerobic capacity or BMI is higher than a benchmark figure determined for each student's age, weight, and gender combination, a student may be assigned to healthy fitness zone status.

[25]We exclude curl-ups, push-ups, and sit and reach from our analysis because they are not completed by a large proportion of the student body.

[26]In the test each time students hear a timed electronic beep they have a set amount of time to run 20 meters (from one line to the other). The exercise ends for a student the second time she cannot finish the 20 meters within the set amount of time. At the end of each minute students hear 3 beeps letting them know that the amount of time they will have to finish the 20 meters has been reduced. A student's score is the number of laps she completed before her second failure to complete the 20 meters within the allotted time. Some schools actually use a one-mile run test to assess aerobic capacity. We do not observe the test employed, however both tests are converted to a comparable scale of VO_2 max. See Boiarskaia et al. 2011 for additional information on how these two tests are converted to the same measurement of VO_2 max, and Blasingame (2012), which finds that both assessment types accurately capture VO_2 max and are consistent with each other.

[27]Given age and weight, the number of laps completed by a student can be used to determine the student's maximal aerobic capacity, or VO_2 max. The Cooper Institute approximates this value based on a functional transformation of the number of laps completed and the student's age. For more information, see the Cooper Institute FitnessGram

also directly measures each student's BMI, which is defined as a student's mass in kilograms divided by her height in meters squared. The CDC defines healthy and unhealthy levels of BMI for children based on their percentile rank among all children of a given age and sex.

The first and second years of FitnessGram aerobic capacity information collected by the state, 2011-12 and 2012-13, are not consistent with the remaining years.[28] These early years feature many average VO_2 max values that are simply not observed in later years. More troublingly, some of these very low average VO_2 values correspond to unrealistically high levels of healthy fitness zone attainment. The indiscrepancies may result from a few possible factors, although we cannot diagnose the precise origin of the issue.[29] We believe the unreliable observations are primarily an issue of accidental half-counting by coaches administering the test for the first or second time. This view is consistent with the findings of Blasingame (2012) that differences between one-mile run and the PACER test and between FitnessGram versions 8 and 9 are minimal. To account for this issue while preserving as much data as possible, we take a rule-based approach to identifying schools that most likely have contaminated scores, dropping any school-level observations below the minimum score by gender that we observe across all years in which we are confident of the data (those after the 2012-13 school year). In section Table B.22 we explore the robustness of our results to a wide variety of alternative methods for dealing with this issue, including dropping the 2011-12 school year entirely, confirming that our main results are indeed quite conservative.

The first panel of Table B.13 presents summary statistics of the FitnessGram tests for aerobic capacity (AC) and body mass index (BMI) aggregated to the district level. We take the district

Reference Guide.

[28]Figure A.17 shows the extent of inexplicable values in 2011-12 and 2012-13, showing how the otherwise tight linear relationship between percent of students in the healthy fitness zone and the average VO_2 max, which we see in the 2014-2017 data, is dramatically less reliable in the first two years.

[29]One potential cause is that schools calculated VO_2 max using the FitnessGram version 8 equation in 2011-12 and 2012-13, whereas in later years they use the conversion equation from FitnessGram version 9. Second, roughly one third of schools implement a one-mile run test while the remaining schools use the PACER test. Although both have been converted to units of VO_2 Max in our data, the correlation between VO_2 max and performance on the one-mile walk is slightly lower. Blasingame (2012) finds that the one-mile run is less correlated to actual VO_2 max than the PACER test (correlation coefficients of .84 and .93), but both assessment types and estimation equations are consistent and generally accurate. Third, coaches may have half counted PACER laps, effectively counting a "down and back" as one lap rather than two. We suspect this issue because more-recent official coaching instructions specifically advise against this counting practice.

average as our outcome measure because treatment in our data is at the district level. Average values were converted from school- to district-level by calculating the sum of weighted school averages for each district, where the weight is the proportion of a district's attempts taken at that school.[30] The attempts divided by enrollment is an approximation of the proportion of students completing a FitnessGram examination in each district. AC and BMI were the two most common FitnessGram examinations, though less than half of students in a district completed the AC exam, while about two-thirds of students completed a BMI examination in any given year.[31]

3.2 Academic Achievement

Our second source of data includes information on student test scores, enrollment levels, and demographics from the Georgia Department of Education (GADOE), which provides school-level data from 2006-07 to 2016-17. Only English language arts (ELA) and math end-of-grade 3rd-grade through 8th-grade test scores are reported throughout the sample window, so we focus on these exams. The recorded information includes the average raw scale score of students in each grade and the number of student test takers for each test. We normalize scale scores using the state mean and student-level standard deviation, and then average over grades and schools using weights for the number of test-takers. This yields a district-level average performance, in terms of student-level z-scores, for ELA and math in each year of the sample. From 2013-14 to 2014-15, the state changed its assessment regime from the Criterion-Referenced Competency Test (CRCT) to the Georgia Milestones Assessment System, with an accompanying change in scale and difficulty on

[30]For example, district i's average aerobic capacity in a given year is $y_{it} = \sum_{s=1}^{N} x_{st} \frac{a_{st}}{a_{it}}$ where x_{st} is the school average in year t and a is the total attempts on the relevant FitnessGram examination for each school s in district i and year t. Alternatively, the weights could be school-level and district-level enrollment instead of total attempts, but this aggregation procedure overemphasizes schools that have lower levels of FitnessGram participation, such as high schools.

[31]Some students are not tested because children below 3rd grade do not take the test, and any students who are not in a physical education class also do not take the test. Additionally, tests administered to fewer than 25 students in a school are censored to protect privacy, hence some school observations are missing. In Table B.23, we find no relationship between FitnessGram attempts on aerobic capacity and bus retrofits. It is also impossible to know whether the total attempts reported by the state reflect multiple attempts by the same student. This could introduce noise if, for example, districts compensate for lower performance by allowing their students more attempts, which would tend to mute physical fitness differences across districts. We also test for this possibility in Table B.23.

the math end-of-grade exam. This is accounted for by normalizing within grade-year and including year fixed effects in our regression models.[32] The second and third panels of Table B.13 report district-level schooling outcomes and demographic characteristics. Test scores are slightly higher for retrofitting districts,[33] though this may be confounded by the effects of the retrofits themselves. Attendance rates are virtually identical across retrofitting and non-retrofitting districts. On average, non-retrofitting districts are smaller, but have otherwise similar student compositions.

3.3 Bus retrofits

The third data source contains information on all bus retrofits from 2003-2018 and was provided through an open records request by the Georgia Environmental Protection Division (EPD). These data describe the type of retrofit performed in each district, the number of buses affected, the month and year of implementation, and the specific grant used to finance the retrofit. We use district-specific invoices for reimbursement for installation of retrofits to calculate the amount each district paid for their retrofits.[34] Figure A.15 maps retrofitting districts. The fourth panel of Table B.13 shows that a typical retrofitting district improved 66 buses, or close to 19% of the bus fleet, in each retrofit cycle.

3.4 Bus manifest

We augment this with the Georgia Transportation Authority's manifest of all state school buses from 2010-2016. Since the bus manifest covers fewer years than for which there exist retrofits, information for 2007-2010 and 2017 is replaced with the value of the nearest available year in the

[32] Later, in Table B.24, we drop the Milestones years from the sample. Aside from being a slightly different examination, there were widespread issues with the new computer-based assessment. The state notably decided not to use the Milestones examination for accountability purposes in 2015 and 2016.

[33] Standardized test score averages are different from zero because there are many low-performing districts with small student populations and a few high-performing districts with many students.

[34] Although we do not observe actual emissions pre- or post-retrofit, the EPD does provide predictions of the yearly and lifetime reductions of four pollutants using the EPA Diesel Emissions Quantifier. These pollutants were fine particulate matter (PM2.5), volatile hydrocarbons, carbon monoxide, and nitrogen oxides. Because these are predicted emissions changes based on engineering models rather than measured or observed values, we do not use these data.

sample.[35] The manifest includes specific bus identifiers, type of bus, capacity, and bus manufacturing details like make, model and year, fuel source, passengers, daily miles, and the number of students living within 1.5 miles of the school who are eligible to be riders. Some of these statistics are summarized in the last panel of Table B.13. The variety of information provided by the bus manifest allows the creation of variables for the district-wide average student minutes spent in the bus, the district-wide bus ridership rate, and the proportion of district buses retrofitted for each grant. These three variables comprise our treatment measures. In our sample, the average bus rider spends a little less than 45 minutes on the bus each day. The average district has a 62% bus ridership.

4 Empirical Strategy

Our identification strategy exploits variation in the timing and location of retrofits across Georgia. We adopt a first-differences estimation strategy, which differences out any unobserved, time-invariant district attributes that might be correlated with retrofits and health or achievement. The estimating equation is as follows:

$$\Delta y_{it} = \beta R_{it} + \Delta X_{it}\gamma + \tau_t + \Delta\varepsilon_{it}. \tag{2.1}$$

All variables are aggregated to the district (i) year (t) level as described above. Δ indicates a one-period change in a variable (e.g., $\Delta y_{it} = y_{it} - y_{it-1}$). The dependent variable, y_{it}, can be either one of the two health outcomes (aerobic capacity and body mass index) or one of the three schooling outcomes (math and English scores and attendance). Since many retrofitting districts experience more than one retrofitting episode, the model captures these year-on-year changes in health and schooling as a result of proportional changes in the share of buses retrofitted.

Our treatment variable, measuring district retrofits that occurred between time $t-1$ and t, is R_{it}

[35] Inclusion or exclusion of these years does not affect the sign or diminish the magnitude of the results, as we show in Table B.25.

(one can think of R_{it} as the change in cumulative retrofits between $t-1$ and t).[36] We consider three different ways of measuring treatment intensity, R_{it}. The first measure is the proportion of the bus fleet retrofitted that year, termed *Percent Retrofitted*. For example, if a district retrofits 10% of its buses between $t-1$ and t, then $R_{it} = 0.1$. In this case, the magnitude of the coefficient on R_{it} shows the effect of retrofitting an entire fleet – going from all dirty buses to all clean buses.[37] The second measure is the proportion of the bus fleet retrofitted multiplied by the time-constant proportion of students in the district who are bus riders, termed *Percent Retrofitted * Ridership*. For example, if 10% of buses were retrofitted between time $t-1$ and t, and time-constant average bus ridership in district i is 50% of students, then $R_{it} = 0.05$. Here, the coefficient on R_{it} shows the effect of retrofitting an entire fleet in a district where all students ride the bus. This accounts for the fact that the impact of retrofitting should have a larger effect in districts where a higher fraction of students ride the bus. We use time-constant district averages for the proportion of students who are bus riders to avoid identifying changes off potentially endogenous ridership changes, though we later show that changes are in fact small and unrelated to retrofits. Our third measure is the proportion of the bus fleet retrofitted times the fraction of students who are bus riders times the time-constant average duration of each bus ride in minutes per day. This is termed *Percent Retrofitted * Ridership * Trip Duration*. Here again we use the time-constant district average for bus ride minutes to avoid identifying effects off potentially endogenous changes in trip duration. Given two district-years with an equal proportion of buses retrofitted and an equal share of students who ride the bus, if one district buses students twice as far as the other, we should expect larger effects in that district.

Equation 2.1 includes the vector ΔX_{it}, measuring annual changes in the following district-level student characteristics: percent of the student body that is Asian, Hispanic, African-American, male, English-language learner, eligible for free- and reduced-price lunch, or possessing of a disability. The vector ΔX_{it} also includes the following district-level changes in bus fleet characteristics: average bus age, to account for new buses replacing older models, the share of buses that are

[36] In other words, we could also have modeled this as ΔR_{it}^{cumul}, the change in cumulative retrofits. This causes difficulties when we interact R with the share of students who are bus riders because we do not want to identify variation resulting from potentially endogenous changes in ridership.

[37] The average proportion of the fleet retrofitted for the observed retrofits is 0.189.

older models made before recent emissions regulations, student ridership, trip duration, and the share of buses that run on liquid natural gas, regular gasoline, and butane. We find little impact from their inclusion. τ_t is an academic year fixed effect.

Our identifying source of variation is the timing and magnitude of the retrofits. Differences in the share of students riding the bus and the average length of ride among riders add additional variation. An identifying assumption is that this timing is uncorrelated with any potential confounders that would affect health or academic performance. This assumption would be violated if, for example, retrofit timing was a function of expected changes in health or academic performance. Such endogeneity is unlikely in practice because funding allocation decisions were made by a state agency, the Environmental Protection Division, independently of any school district prerogatives. Moreover, the timing of bus retrofit completion varied greatly within grant cycles and across districts. Still, if this were true, we would also see changes in BMI as a proximate health outcome, which we test for. We might also be concerned with endogenous responses on the part of students and families through, for example, increased ridership in response to cleaner retrofitted buses. We employ several robustness tests to allay each of these concerns and discuss each in turn directly following our main results.

One could also estimate the model using district fixed effects. This requires stronger assumptions than the first differences model, some that we likely do not satisfy. For example, the first differences model best captures immediate year-on-year changes, given that a large number of districts have multiple retrofit cycles. More importantly, we worry about serial correlation. First differences requires only that R_{it} is uncorrelated with $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$ where fixed effects requires R_{it} to be uncorrelated with $\varepsilon_{it} - \overline{\varepsilon_i}$ (i.e., that all errors are uncorrelated as opposed to uncorrelated changes in errors, which is a weaker assumption). In the absence of serial correlation, the fixed effects estimator has consistency advantages over first-differences, but as we show in robustness checks, Durbin-Watson statistics suggest that we do not satisfy this requirement, in particular for academic outcomes which are highly serially correlated. Moreover, we have a relatively large number of time periods (10) compared to the number of individual observations (180), which again leads to advantages in the first differences model as fixed effects assumes $N \to \infty$ with fixed T. Regardless, as part of our many robustness tests we also report estimates from fixed effects regressions. We show that while point estimates are similar in nearly all cases, standard errors are larger under the fixed effects model, which is consistent with our concerns.

5 Results

5.1 Health

We present our main regression results for aerobic capacity and our placebo health outcome, BMI, across all three measures of treatment R_{it} in Table B.14. These regressions are based on Equation 2.1 and use data from 2012-2017. The first three columns present effects on aerobic capacity (AC), where the units represent VO_2 max, which is measured in milliliters of oxygen intake per kilogram minute. The second three columns present the effects on BMI. The coefficient in column 1 implies that if a district retrofitted 100% of its fleet, average VO_2 max would increase by 1.8 units, or about a 4% increase relative to the baseline mean of 41.16. Since the average retrofit affected 19% of the bus fleet, the average retrofit improved district-wide aerobic capacity by 0.33 milliliters of oxygen per kilogram minute.

Columns 2 and 3 use the alternate measures of the treatment effect R_{it}. In column 2 it is the percent of buses retrofitted multiplied by the percent of students who ride the bus. This coefficient implies that if a district had 100 percent ridership *and* retrofitted its entire bus fleet, average student aerobic capacity would increase by 2.4 units, or about 6 percent of the mean. The average bus ridership rate is 62%, so this implies that the average retrofit (19% of the fleet) in the average district increases aerobic capacity by 0.28 milliliters of oxygen per kilogram minute. Column 3 sets R_{it} to the percent of the bus fleet retrofitted times the ridership rate times the average trip duration. The coefficient implies that, if all buses in a district are retrofitted and all students ride the bus, then each additional minute of bus riding for students in this district is associated with roughly 0.041 units increase in VO_2 max. Since the average trip duration is 46 minutes, this implies that the average retrofit in the average district increased aerobic capacity by 0.21 units VO_2 max.[38] Thus our point estimates, when scaled, are roughly consistent across specifications in the range of 0.2 to 0.4 units VO_2 max. Given that there is little variation across retrofitting districts in the ridership share and trip length, we do not find this result surprising.

We next turn to our placebo health outcome, BMI. In the final three columns of Table B.14 we find that estimates are effectively zero in all cases. Although directions suggest lower BMI, the coefficient on our main estimate (-0.24) is equal to approximately 1% of BMI. We take this as suggestive evidence that retrofits were uncorrelated with general health trends across treatment and control districts.

We next break out results by gender and school level. Table B.15 displays male and female aerobic capacity results in the full sample across elementary, middle, and high schools. These results reveal two pieces of information. First, estimates are comparable for male and female students. While point estimates are different across gender for elementary school students, the coefficients for male and female students at a given level are not statistically different from one another. Second, effects are highest among elementary school students. We find noisy and in fact negative effects for boys in middle school. Although we are unable to explain this, we believe it relates to influence of outliers in the middle-school assessments and the likely re-assessment of physical education classes to selectively lower-quality students after one mandatory year of the course. The consistency across elementary male and female estimates contradicts the hypothesis that differential incidence of childhood asthma in young boys would exert some influence on these relative effect sizes (Bjornson and Mitchell, 2000). As has been shown in other work (Beatty and Shimshack, 2011), children with asthma are more susceptible to the negative effects of air pollution.

[38] $= 0.189 * 0.62 * 46 * 0.041$

5.2 Academic Achievement

We present our main regression results on three academic outcomes in Table B.16. These regressions include years 2007-2017, since we observe test scores for more years than we observe FitnessGram outcome measures. In columns 1-3, the dependent variable is a z-score of average English (ELA) test scores, normalized to the student-level standard deviation, for grades 3-8. The coefficient in column 1 implies that retrofitting an entire fleet would raise ELA scores by 0.09 standard deviations. This represents an achievement differential slightly larger than that observed between students of a rookie teacher and those of a teacher with five years of experience.[39] The average retrofitting district retrofitted 19% of the fleet, suggesting an average increase in ELA scores of 0.017 standard deviations per retrofit cycle. In column 2, the treatment effect R_{it} is the share of buses retrofitted times the share of students who ride the bus. The point estimate suggests that retrofitting an entire bus fleet with 100% ridership would increase student test scores by 0.145 standard deviations. The average retrofit (19% of the bus fleet) for the average district (61% ridership) increases scores by 0.017 standard deviations according to this point estimate, which is identical to the result in column 1. Column 3 shows that each minute of bus riding in a 100%-retrofitting district with 100% ridership is associated with a 0.003 standard deviation increase in ELA scores. Based on this, the average district's retrofit increases ELA scores by 0.016 standard deviations, which is consistent with specifications (1) and (2).

The results on math test scores (columns 4-6) are also positive but only about one-half as large as the ELA results and not statistically distinguishable from zero. This is consistent with Ham et al. (2014) who find that particulate matter, and especially PM2.5, tends to affect ELA scores more than math scores. Specifically, they find that PM2.5 lowers math scores by 60% less than ELA scores, which is similar to our findings. The last three columns of Table B.16 show that there is no effect of retrofits on average attendance rates. Since the mean attendance rate is 0.95, there is little margin for gain. This contrasts with the negative attendance effects found in Adar et al. (2015). In Table B.17, we show how the percentage of a bus fleet retrofitted affects ELA and math

[39] Rice (2010).

z-scores among elementary and middle school students.[40] Consistent with the health estimates, effects are larger in elementary schools than in middle schools. For both elementary and middle schools, the effects on math are positive but indistinguishable from zero.[41]

5.3 Results by Retrofit Type

In Table B.18, we present results by type of retrofit for each of our academic and health outcomes. There were few episodes of closed-crankcase filter retrofits (244), and fewer of flow-through filter retrofits (58). In fact, there were none of these retrofits over the sample period during which we observe aerobic capacity and BMI records from the FitnessGram examination. Nevertheless, diesel particulate filters (1,160 retrofits) and diesel oxidation catalysts (1,394 retrofits) had a positive and roughly consistent effect on both ELA and math test scores. Adar et al. (2015) find that implementing DOCs and CCFs both had an effect on attendance, with larger and more significant effects for DOCs. This is consistent with our findings for DOCs only, the discrepancy likely caused by the low number of CCF retrofits. We add to their findings by testing for effects on diesel particulate filters, which appear to have a larger effect on ELA, math, and aerobic capacity. Since DPFs are expected to eliminate 60-90% of fine particulate matter, while DOCs eliminate 10-50% of fine particulate matter in Adar et al. (2015), this finding appears reasonable.

5.4 Robustness and Alternate Specifications

Alternate Specifications

We re-estimate our main results (both health outcomes and academic outcomes) using a fixed effects specification. In the top panel we show our main results from the baseline first-differences

[40] We do not have test scores by gender, nor do we have them for high school students.

[41] In Table B.20 we display results dis-aggregated by grade. The grade-level performances are consistently in the same direction as the main academic estimates, and achieve significance in at least one grade for each ELA and math test scores. Interestingly, grade-level effects suggest larger impacts for students more likely to sit at the back of the bus– those in 4th, 5th, and 8th grade– which is consistent with bus self-pollution from diesel exhaust.

model for reference. In Panel II we show estimates from a fixed effects model. This is:

$$y_{it} = \alpha + \beta R_{it} + X_{it}'\gamma + \tau_t + \phi_i + \varepsilon_{it} \tag{2.2}$$

Where ϕ_i is a district fixed effect. Point estimates for all outcomes are similar across Panels I and II, though standard errors are larger under fixed effects. The exception is the coefficient on math scores, which becomes zero under fixed effects. Panel II also show Durbin-Watson test statistics, which indicate a high degree of serial correlation. Given that the math estimates were zero in our main specifications, a smaller coefficient here does not change conclusions. For ELA, we find a marginally larger point estimate and substantially larger standard errors, marginally failing to reject the null. Effects on aerobic capacity are similar though noisier as well. Taken together, these two panels of Table B.19 suggest efficiency gains from first differences as expected in the presence of serial correlation, and that our conclusions are not substantively altered by our modeling choice.

We also use the fixed-effects model to conduct additional robustness checks. We re-estimate Equation 2.2 above and include a lead of R_{it} ($R_{i,t+1}$) as a test for pre-trends. Panel III of Table B.21 shows the results. Academic results are similar to our main estimates (the ELA coefficient is slightly larger), and we cannot reject that the lead coefficients are zero for each outcome except aerobic capacity. We find that the lead for aerobic capacity is large and significant. We believe this is strongly influenced by the noisy first year of AC data. When we add district time trends to the fixed effects model with leads, in panel IV, the lead coefficient in the AC regressions (column 4) is not statistically significant and the contemporaneous coefficient returns to the same magnitude as in the base case.

In this model, the coefficient on ELA test scores shrinks from 0.089 in the first-differences specification to 0.032.[42] These results suggest that the fixed effects estimates are sensitive to modelling decisions. We note that 9 of 10 lead tests fail to reject the null hypothesis of no leads,

[42] Our preferred first-differences specification is robust to the inclusion of leads and district-specific trends. In Table B.21, we include two leads and district trends in all regressions. For the FitnessGram test regressions (columns 4 and 5), we cannot include more than one lead since there is not enough variation in the retrofits during the years we have the FitnessGram test data. But, inclusion of the second lead does not substantially change our results.

supporting the notion that we are not picking up spurious correlations. In Table B.26 below, we conduct a stricter test of timing by assigning treatment to one year in advance, and find that all estimates are zero, as one would expect if pre-trends are not a concern. Similarly, in the next section we show pre-trend plots. We expect that the trend and lead relationship we observe with respect to aerobic capacity is related to outlier data in the first year of our sample, which we address in section 5.4 to follow.

Academic Achievement Pre-Trends in Retrofitting Districts

One might be concerned that retrofitting districts have different pre-treatment trends that drive the results. This possibility is difficult to test directly because there is no uniform year of treatment across retrofitting districts. For this reason, there are no uniform pre-treatment or post-treatment years. Many retrofitting districts also had multiple retrofit cycles. To assess the possibility of differential pre-trends, we therefore plot academic achievement outcomes from 2006-07 to 2011-12 across retrofitting and non-retrofitting districts in Figure A.16. We plot results in the years leading up to 2013 because this was the modal retrofit year with nine retrofits. We note that 25 retrofits occur before this year, so we may expect the slope trends to increasingly differ by the extent to which the retrofits impact academic outcomes. Nevertheless, the trends appear close to parallel over this period. We do not plot pre-trends for our health outcomes because of the shorter window over which we observe these outcomes and the notable issues with aerobic capacity information in the roll-out year of the program (as discussed in section Table B.22).

Aerobic Capacity Data

As discussed earlier, the early FitnessGram results contain inconsistencies, so we apply a rule-based approach in which we eliminate implausible values. In Table B.22 we re-estimate our main specification, using the share of buses retrofitted, across different cutoff values to demonstrate how our results vary across different rules of thumb. The first five columns of the table show results for cutoffs set at $15, 20, 25, 30, 35$. These represent dropping school-level aerobic capacity results below the given value in 2011-12 and 2012-13 (although, in practice, almost all removed values are in 2011-12). In column 6 we show our preferred cutoff of 26 for females and 30 for males for reference, the lowest observed values after 2012-13. In column 7 we apply an alternate rule where we eliminate schools for which we observe a jump of more than 6 in Aerobic Capacity – equivalent to 15 percent of the mean – between 2011-12 and 2012-13 as an indicator of reporting issues in the first year. In column 8 we show the full data, not dropping any schools, and in column 9 we show effects if we drop school year 2011-12 entirely. With the exception of the specification in columns 7 and 9, results are similar in magnitude across specifications. Eliminating problematically low observations affects the standard errors, as we would expect. In column 7, when we drop implausibly large jumps, estimates double, and when we drop the first year of data entirely in column 9, effect sizes increase over four-fold, from 1.8 to 7.1. While we are more confident in these estimates, we take the conservative case of only dropping problematic observations as our preferred estimate.

Correlation of Proportion of a Bus Fleet Retrofitted with District Characteristics 2007-2017

We address the potential for retrofits to affect participation in the FitnessGram test, possibly due to increased health status, in the first panel of Table B.23. In columns 1 and 2, we regress the participation rates for aerobic capacity and BMI FitnessGram tests, measured as the total number of test attempts divided by the district enrollment, on the percent of a bus fleet retrofitted. We find no discernible relationship between district retrofits and the share of student who are tested in aerobic capacity. If anything the point estimate suggests a small negative relationship. We find a similar pattern for BMI tests, suggesting that districts with more retrofits see a marginally higher rate of BMI testing, though again the estimate is noisy. In column 3 we test for changes in ridership, potentially resulting from an increase in the share or number of students riding the bus as a result of reduced emissions. We find no effect, suggesting that cleaner buses do not increase ridership. In the same table, we demonstrate the relationship between the proportion of a bus fleet retrofitted and changes in bus fleet characteristics, student demographics, and student characteristics. We observe a statistically significant relationship in only one case; the proportion of students with disabilities in a district is positively related to the proportion of a bus fleet retrofitted. We believe it is unlikely that retrofits would change disability status among students. Regardless, we control for changes in the share of students with disabilities in all regressions.

Milestones Test Sensitivity

The roll-out of a new Milestones exam (Georgia's end-of-year test) in 2015 resulted in a large decreases in math scores in several of Georgia's largest districts, many of which received retrofits. The decrease was caused by complications in the new internet-based math examination where several districts had computers "freeze," causing severe disruption to test-takers.[43] As a result, those exams were not used to calculate district performance for state requirements, student retention, or graduation.[44] That retrofitting districts are primarily Georgia's larger districts, which were those who adapted to computer based tests, raises concerns that this could be a confounding factor in our test score analysis. When we drop Milestones years 2015-2017 from the sample, in Table B.24, the results are qualitatively similar to our main specification, although math scores are larger in magnitude and more precise. Since no districts retrofitted after 2015, this change is not correlated with contemporaneous treatment, but rather shows a decline in test score post-treatment for these districts.

Exclusion of Interpolated Bus Manifest Data

The district bus manifest covers 2009-10 to 2015-16. We fill in the remaining years by substituting the value of the nearest chronological neighbor for each year. For example, a district's 2016-17 value for total buses is set equal to the number of buses it had in 2015-16. Linear interpolation was ruled out because it created unrealistic values for some districts with large changes in their bus fleet. As shown in Table B.25, our results are unchanged by the exclusion of years for which we lack information on district bus fleets. In fact, excluding these years improves the precision of

[43] Cobb, Dekalb, Cherokee, and Gwinnett counties all suffered from these computer glitches.

[44] See this article and this article for more information.

both our math and ELA point estimates.

Timing of Retrofit Treatment

There are two sources of imprecision with respect to the timing of treatment. First, the FitnessGram test may be in fall, spring, or both, while the end-of-grade tests are uniformly in April-May.[45] Second, the date of the bus retrofit reimbursement invoice, which we use as a proxy for the date of retrofit completion, imperfectly corresponds to the date when the buses are first used. If the timing of a retrofit comes before April of the year in question, the retrofit is counted as occurring in that school year even if some of the FitnessGram tests may have occurred before the retrofitted buses were active. This may affect the results of some FitnessGram tests while leaving the test score results unaffected. On the other hand, buses completed in a retrofit before April may not actually be used until the following school year due to implementation lags, which would mean our baseline treatment year assignment is too early to pick up changes in test scores. In Table B.26 we show our baseline treatment assignment and explore a placebo timing treatment that assigns the year of the retrofit to one year in advance of the year of the retrofit completion invoice. These results, presented in the second panel of Table B.26, demonstrate that the assigned treatment timing is not inconsequential, as no estimate is significant when adopting a placebo treatment year. In Panel III we assess the possibility that our treatment assignment for retrofits occurring after January is too early by assigning the same fiscal year to any retrofits completed before January and the subsequent fiscal year to any retrofits completed after January. Under this treatment year assignment rule, the results are the same for each outcome except for math test scores, which are now positive and significant. We take this as suggestive evidence that our baseline treatment assignment is not too late to capture changes in aerobic capacity, although it may be too early to pick up changes in academic achievement for some districts.

[45] Across the state we know that two-thirds of FitnessGram exams are given in Spring and one-third in Fall, although we do not know the breakdown by district.

Linear Trends

In Table B.27 and Table B.28, we present results with the first-difference model including linear trends. Adding district-specific linear trends amounts to adding a district fixed effect in the regressions (first-differencing eliminates a district-fixed effect and converts a linear trend to a fixed effect). Table B.27 replicates the health regressions (Table B.14), and Table B.28 replicates the academic regressions (Table B.16). Both sets of results are robust to inclusion of trends. For health outcomes, adding linear trends slightly increases the magnitude of the effect of each definition of treatment on aerobic capacity. For academic outcomes, adding linear trends slightly reduces estimates for ELA scores, although conclusions are similar.

6 Cost-Benefit and Cost Effectiveness Analyses

We conduct back-of-the-envelope calculations of the costs and the benefits of bus retrofits. We examine health benefits in terms of both reduced mortality and reduced cardiovascular disease, as well as benefits from increased test scores. We note that this does not account for spillover effects on non-treated members of the community who are exposed to lower pollution levels overall. Additionally, we compare the cost of achieving the education benefits from the retrofits to the costs of achieving similar gains from class-size reduction to provide a cost effectiveness analysis.

6.1 Costs

The total amount awarded for district bus fleet retrofits in Georgia is \$26 million. However, certain retrofits occurred before our sample window. Moreover, a large portion of funds went to purchasing new buses to replace older ones. We separate the amount awarded for bus replacement from the amount spent on retrofits using invoices detailing each district's reimbursement for completing their retrofit. These reimbursements include the cost of parts, labor, and daily usage of a repair bay. The total amount spent on engine retrofits is \$12.6 million, with the average district spending \$8,110 per retrofitted bus. The average district has 111 buses, so the cost of the average district retrofitting 10% of its fleet is \$90,000. For comparison, the cost of one regular new bus is roughly \$130,000, while a new hybrid or electric bus is \$360,000. Replacing 10% of a fleet with new diesel or hybrid buses would therefore cost \$1.4M - \$4M, an order of magnitude greater than the cost of engine retrofits.

6.2 Benefits - Health

We focus on the health benefits in terms of increased aerobic capacity, which is the most persistent result. Our preferred specification is column 1 of Table B.14, which indicates that a ten-percentage-point increase in the percentage of buses retrofitted is correlated with a 0.18-unit increase in the measure of aerobic capacity. The units we observe for the aerobic capacity measure are milliliters oxygen per kilogram minute (mL/min/kg); these units have already been converted into a measure of VO_2 max from the number of PACER laps completed using a standard conversion factor provided by the FitnessGram test manufacturer.

From this conversion we conclude that a ten-percentage-point increase in the percentage of buses retrofitted is correlated with a 0.18-unit increase in VO_2 max. We convert the VO_2 max effect measure from units of mL/min/kg to units of metabolic equivalent (MET) by dividing the VO_2 max in mL/min/kg by 3.5, yielding a change in MET of 0.05 for a retrofit of approximately 10 percent of a district's bus fleet.[46]

Several studies document and measure the benefits from increased aerobic capacity (or cardiorespiratory fitness).[47] Kodama et al. (2009) conducts a meta-analysis and finds that a 1-MET higher level of VO_2 max is associated with a 13% decrease in the risk of all-cause mortality and a 15% decrease in the risk of cardiovascular disease (CVD).[48] However, this meta-analysis was conducted on studies of adults, not children. Other studies examine the effect of cardiorespiratory fitness on children's CVD outcomes[49], but do not provide an estimated magnitude of a causal effect

[46] Castillo-Garzón et al. (2006).
[47] Several such studies are summarized in Institute of Medicine (2012), Chapter 5.
[48] Lakoski et al. (2015) finds also an association between aerobic capacity and adult cancer rates.
[49] Castro-Piñero et al. (2017), Ortega et al. (2008)

from VO_2 max.

We thus use two different measures of the valuation of health benefits from aerobic capacity increases. First, we use the meta-analysis of mortality effects reported in Kodama et al. (2009) for adults and extend them to childhood mortality: a 1-MET increase in VO_2 max is associated with a 13% decrease in mortality risk. The baseline childhood mortality rate in Georgia among 5-12 year olds was 13.3 deaths per 100,000 population in 2016.[50] We use a standard value of a statistical life (VSL) of $7.4 million.[51] The average district in Georgia has about 9,000 students. Thus, if an average district's average MET unit of VO_2 max increased by 0.05 units (the effect size 1.8 scaled to represent a district retrofitting 10 percent of a its buses and divided by 3.5 to convert to MET units), the health valuation from reduced mortality for that district is $71.1.[52] Assuming a retrofit life of 10 years[53] and an annual discount rate of 3%, the present discounted value of the mortality reduction benefits is $624.69, a small fraction of the cost of retrofitting 10% of the bus fleet calculated earlier, $90,000. It is perhaps not surprising that the retrofits fail a cost-benefit analysis when the benefits are calculated only from reductions in mortality, since the baseline mortality rate for elementary-school-aged children is extremely low.

The second measure of the valuation of health benefits combines the result from Kodama et al. (2009) on the effect of aerobic capacity on cardiovascular disease (among adults) with results from Adamowicz et al. (2014) on the valuation of avoided CVD among children. Adamowicz et al. (2014) conduct a stated-preference survey of parents asking for their willingness-to-pay (WTP) for a reduction in the probability of their children being diagnosed with heart disease by age 75. They report a mean annual WTP to reduce that probability by one chance in one hundred of $5.62

[50] https://oasis.state.ga.us/oasis/webquery/qryMortality.aspx#

[51] https://www.epa.gov/environmental-economics/mortality-risk-valuation#whatvalue

[52] The 0.05 MET increase = 0.00000665 PP decrease in the mortality rate = 0.00000960555 averted deaths per average district retrofit
= $71.1 per district.

[53] Diesel particulate filters are often given a lifespan of 100,000 miles by the manufacturer, which represents 8 years with our sample's average yearly mileage of 12,960. However, DPF lifespan varies greatly depending on regular servicing and cleaning. Barone et al. (2010) show that DPFs are 95% as effective after four years, while Sappok et al. (2009) show that DPFs are half as effective at 188,000 miles, or roughly 14 years for the buses in our sample. We select 10 years to be consistent with Beatty and Shimshack (2011), although the entire range (4-14 years) of possible lifespans lead to benefits far less than the costs of $90,000.

for mothers and \$4.08 for fathers; we use the mean of these two values (\$4.85). Since this is an annual WTP, we interpret the total WTP for the one-in-one-hundred chance reduction in CVD to be the net present value of this annual WTP from age 11 until age 75, which equals \$139.34.[54] Kodama et al. (2009) report a 1-MET increase in VO_2 max is associated with a 15% decrease in the risk of CVD. About one third of Americans have some form of CVD,[55] so a 15% decrease in the risk is equivalent to a decrease in the chance of 1 out of 20. Therefore, the benefit from a district retrofitting 10% of its buses is valued at \$940,590 per district.[56] This is more than nine times greater than the cost of the retrofits. Because CVD is so prevalent (unlike childhood mortality), the valuation of even a modest reduction in its risk is quite high. These benefits do not take into account the value of lower pollution levels for non-students.

6.3 Benefits - Test Scores

Next, we calculate the benefit of the retrofits from a monetization of test score improvements. Chetty et al. (2011) estimate the effect of an increase in kindergarten test scores on adult earnings; they report that a one-percentile increase in test scores is associated with an increase of \$94 in wage earnings at age 27 after controlling for parental characteristics. Assume that the wage benefit of \$94 lasts throughout one's working years of age 25-54, and discount using an annual rate of 3%. Then, the one percentile increase in test scores is valued at \$1,041.[57] The results presented in Table B.16 indicate that retrofitting 10% of a district's fleet will increase the z-score of the ELA tests by 0.009 and of the math tests by 0.005. These improvements in z-scores are equivalent to percentile increases of 0.36 and 0.19, respectively. Using the average of these two values (0.275), and multiplying by the valuation implied by the Chetty et al. (2011) estimates, the benefit of retrofitting 10% of a district's fleet is valued at \$2.57 million.[58] This is over 25 times greater than the costs of

[54] The survey sample in Adamowicz et al. (2014) includes just parents with at least one child aged 6-16 in the home, so we use 11 as the starting age.

[55] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5408160/

[56] The 0.05 MET increase = 0.0075% decrease in the probability of CVD = \$104.51 benefit per child = \$940,590 benefit for an average district with 9,000 children.

[57] $= \sum_{i=20}^{50} 94 * (1-0.03)^i$

[58] $=$ 0.275 percentile points $*\$1,041$ per percentile point per student $*9,000$ students per district.

the retrofit.

Lastly, we compare the costs of achieving test score gains through bus retrofits to the costs of achieving those same gains through interventions studied in Chetty et al. (2011). The Tennessee STAR program reduced class sizes by seven students, which is expected to cost around $870 per student,[59] and it yielded a 4.81 average percentile improvement in test scores. Our estimates of the effects of the retrofits are that they yielded a 1.9 - 3.6 average percentile increase in test scores. The average school bus in our sample transports 66 students per day. Since the average cost per bus retrofit in our sample is $8,110, this translates to a cost of roughly $122 per student, or $34.1 - $64.7 per percentile point gain. The cost for an equivalent test score improvement is roughly three to six times higher for the STAR class size reduction than it is for the bus engine retrofits.[60]

7 Conclusion

We estimate the effect of retrofitting diesel school bus engines on student health and academic achievement in the state of Georgia. Retrofits have positive and significant effects on students' aerobic capacity, a measure of respiratory health, but no effect on body mass index, which we take as a placebo. Retrofits also have positive and significant effects on student English test scores, and a smaller and precise effect on math scores. Robustness checks reinforce our findings. Back-of-the-envelope calculations suggest that the benefits of the retrofits were much higher than their costs, and that the academic gains were achieved at a lower cost than they would have been through class size reductions.

This study could be extended several ways. First, use of individual-level data rather than district-level data may improve the precision of the results. Within-district variation in the exposure of students to the retrofits could be utilized if, for instance, individual student health records could be matched with bus routes. This could also allow for determining if treatment effects differ by demographic group. Second, data from other states could be analyzed to test whether the re-

[59] Reichardt (2000).

[60] The class size reduction cost $870 per student for 4.81 percentile gain = $181 per percentile point gain.

sults from Georgia generalize elsewhere. Third, alternative health or academic outcomes could be examined. Linking students to other health outcomes, for example via Medicare data, may provide a valuable measure of health not picked up by FitnessGram scores. With a longer panel, long-term outcomes, including college attendance and labor market outcomes, could be examined. Fourth, we could test the effect of retrofits on outcomes other than health and academic performance such as non-cognitive skills.

Our results have plausible policy relevance. While bus retrofit programs are widespread, very little work has examined their effects. Policymakers interested in physical health and academic performance of children can use bus retrofits as another cost-effective policy tool.

Chapter 3

School Nutrition Expenditures, Local Agricultural Revenues, and Farm-to-School Policies

1 Introduction

In recognition of the economic importance of school-provided meals to local businesses, every state has implemented some form of farm-to-school incentive program, with 42,000 schools participating in some form of farm-to-school policy in 2014-15.[1] These policies are diverse in scope and characteristics. They exist at the state, school district, and school level; they also emphasize different elements of the farm-to-school movement. In general, however, these programs consist of three key policies: serving food sourced from local providers, edible school garden activities, and food educational initiatives such as field trips to local farms and food tastings.

In Georgia, meal provision for the 1.8 million public K-12 students represents at least 3% of the state's meal consumption; local, state, and federal nutrition expenditures for these meals amounted

[1] Christensen et al. (2018), USDA (2017). Incentive programs at the state level often consist of recognition and awards for commitment to the movement. For example, Georgia gives the "Golden Radish" award to school districts meeting certain criteria for excellence in commitment to farm-to-school programs. See more here.

to \$13.8 billion from 2001 to 2017.[2] Of all students in the state, 1.1 million received free or reduced price lunch and breakfast, while 1.3 million students were served by schools with some form of farm-to-school policy. Despite the fact that more students receive exposure to a farm-to-school policy than free-and-reduced price lunch or breakfast, relatively fewer studies have investigated how farm-to-school policies affect students or local economies. In particular, no study has assessed the average long-term role that these diverse policies play on local agricultural revenues.

This study asks whether these district policies are associated with changes in local agricultural revenues.[3] I also show evidence on the baseline role that school district nutrition expenditures play on local agricultural revenues. Finally, I provide suggestive evidence as to the types of agricultural products that are most associated both with farm-to-school policies and school district nutrition expenditures in general. To answer these questions, I rely on variation in the timing of farm-to-school policy adoption and the share of students induced into free meal status by the Community Eligibility Provision, which increased school nutrition expenditures. I secure plausibly-causal identification with two statistical models: a triple-difference model relying on school-level adoption of the Community Eligibility Provision (CEP) of the Healthy Hunger Free Kids Act and a two-stage least squares model that uses expenditure changes associated with the CEP to predict local agricultural revenue changes. I find that as much as 7% of school district expenditures flow to producers within the same county. Of this local share, perhaps as much as 70% of the expenditures are associated with adoption of farm-to-school policies. Specifically, \$680M out of \$966M local expenditures may be attributed to farm-to-school policy adoption. These figures represent 0.4% and 0.6%, respectively, of all agricultural revenues in the state from 2001-2017. These local expenditures are more strongly associated with revenue increases of fruits and vegetables than with animal products. This study contributes to a sparse literature on the role that school nutrition expenditures and

[2]One third of meals for 1.8M/10.5M people in the state for 180 out of 365 days in the year (assuming only one meal per student per day, which notably excludes breakfast).

[3]For traction I adopt a definition of local corresponding to the same county and contiguous counties. The definition of "local" varies by program. According to the 2015 farm-to-school census, 27.2% of Georgia farm-to-school programs use a definition of local corresponding to a region within a 100 mile radius, 36% use a definition of local corresponding to the entire state, 18% use a definition of local corresponding to a region of nearby states, and 9% use some other definition of local.

sourcing policies play on local agricultural revenues.

2 Prior Work

A large literature has investigated farm-to-school policies. Joshi et al. (2008) provide an early synopsis of this literature, categorizing farm-to-school studies by those relating to changes in students and parents, those relating to changes in teacher, administration, and cafeteria practices, and those relating to farmer behaviors. For concision, I will focus on the latter category. Many survey-based studies have sought to conceptualize the economic context of farm-to-school policies, examining scalability of farm-to-school policies, institutional factors surrounding the policies, and the barriers to success of the programs; several studies rely on the 2015 USDA Farm to School Census or the USDA Census of Agriculture's 2015 Local Food Marketing Practices Survey.[4] These surveys, however, are cross sections of one year. Moreoever, relatively few studies focus on observable economic impacts of farm-to-school policies rather than school survey-based expenditure information. In part, the paucity of studies directly linking expenditures to local revenues relates to identification hurdles. Selection bias, spatial autocorrelation, policy spillovers, and even reverse causation each contribute to the confounding of correlational estimations.[5]

Christensen et al. (2018) provides an overview of eight studies assessing the economic impacts of farm-to-school programs. Six of the studies were not peer-reviewed, employ varying methods that complicate cross-state comparisons, rely on short (typically one-year) time windows, and often do not use primary data. The authors highlight two case studies of special merit, one in Georgia and the other in Minnesota. In these studies, information on school district expenditures on local food was combined with surveys of farmers supplying to school districts. These measures were inputted into a software known as IMPLAN (IMpact Analysis for PLANing) that allows for

[4] Botkins and Roe (2018), Deller et al. (2017), Holland et al. (2015), Hoffman et al. (2017), Lee et al. (1980),Thompson et al. (2014)

[5] O'Hara and Benson (2019) show that local production conditions are associated with patterns of local-food purchasing by districts, suggesting the presence of reverse causation whereby supply and local production determines the extent to which school districts purchase locally (and not a farm-to-school policies).

disentangling direct, indirect, and induced economic impacts both with and without opportunity costs. The key finding of these studies is that each dollar spent on farm-to-school programs has an implied multiplier effect of 1.5 on output with opportunity costs and 2 without opportunity costs.[6] In a related work, Christensen et al. (2019a) show that farm-to-school purchases from direct local vendors instead of intermediate suppliers is cheaper for school districts.

This study adds to the previous work in several ways. First, I provide estimates of agricultural revenues related to nutrition spending across all farmers and all school districts in the state. The case studies mentioned in Christensen et al. (2018) average information from only seven Georgia farms and five farms near Minnesota. Moreover, whereas food expenditure data in the Georgia case study was based on the stated farm-to-school nutrition expenditures of 61 school districts in one survey year (2014-15), I incorporate all nutrition expenditures across all districts over 17 academic years.[7] This broader reach allows estimation of baseline local food purchases by school districts that do not engage in farm-to-school policies. Next, I observe detailed commodity revenue information, allowing me to separately estimate effects over commodity groups and by specific commodity. Only one study surveyed by this paper investigated how farm-to-school purchases may be broken up by type of commodity, and that study relied on purchase information in a relatively small geographic region during only the first year of a grant-funded policy roll-out.[8] Finally, I incorporate variation in nutrition expenditures associated with a policy unrelated to farm-to-school adoption to secure identification plausibly purged of selection bias and reverse causation.

3 Data

I rely on six sources of information on school districts and one statewide survey of county-level agricultural revenues. Although I observe farm-to-school policies at the district level, I aggregate

[6] The average multipliers in the Minnesota and Georgia studies. These measures closely resemble the spending multipliers found in the six unpublished studies briefly discussed in the paper, which find spending multipliers of 1.1-2.4.

[7] The expenditure data is from the federally-mandating accounting ledgers of school districts, ensuring accurate reporting.

[8] Watson et al. (2018)

to the county level of analysis for estimations on local agricultural revenues.[9] I describe each data source below.

3.1 Farm-to-School Policies

Information on the adoption of farm-to-school policies comes from school district applications for a statewide incentive program known as the Golden Radish Award.[10] A wide range of information on heterogeneous school district farm-to-school policies is included on these applications for the award, which is administered by a non-profit organization known as Georgia Organics in cooperation with the Georgia Departments of Education, Health, and Agriculture. Of special interest on the application is the first year a district implemented farm-to-school policies, which often predates the first application for a Golden Radish Award.[11] The program has expanded dramatically; 30 school districts received some form of Golden Radish recognition in fiscal year (FY) 2014 while over 75 school districts did so in FY 2017. The Golden Radish incentive program appears to affect school district culture, with 40 school districts institutionalizing farm-to-school language in their school district policies.[12] A map depicting all school districts ever adopting some form of farm-to-school policy is presented in Figure A.18. Table B.29 shows certain key summary statistics broken up by farm-to-school policy adoption status. I supplement this information with the 2015 USDA Farm-to-School Census, which reports a wide range of information on county-level nutrition expenditures on farm-to-school foods.[13]

[9] In general, school districts and counties are geographically the same. Some counties, however, have school districts specific to a city within the county. Figure A.15 depicts the overlap of school districts and counties.

[10] The program recognizes districts with five different levels of engagement to farm-to-school policies. These levels are platinum, gold, silver, bronze, and honorary. An additional "outstanding" award is given to one district. The application portal is here.

[11] The survey also includes information on the district-wide number of farm visits, days serving local food, local meals, farm promotions, farm-based classroom lessons, schools with gardens, professional development staff, and local food taste tests among others. Information from these other questions is often missing, available for a relatively short sample window, and self-reported, with possibly heterogeneous definitions and reporting standards across districts. Information on these variables is presented in Table B.36.

[12] Golden Radish Infographic.

[13] USDA (2017).

3.2 Community Eligibility Provision Adoption

Part of the Healthy Hunger-Free Kids Act, the Community Eligibility Provision is a universal free-meal option available to schools or entire school districts with at least 40% of their students qualifying for free lunch through categorical eligibility. The fraction that is categorically eligible, or the Identified Student Percentage (ISP), is the share of students receiving any other form of federal financial assistance. Depending on the ISP, a school district receives federal reimbursement of between 64% and 100% of their nutrition expenditures.[14] The Georgia Department of Education maintains lists of schools and school districts that are eligible for participation in the Community Eligibility Provision, including information on schools or districts that actually participate in these programs from FY 2016-2019.[15] Measures of the number of eligible students, the number of students in participating schools, and the number of marginal students induced into free lunch by the program are depicted in Table B.29. These figures are broken up across districts that ever have or do not ever have a farm-to-school policy.

3.3 Nutrition Expenditures

The Fiscal Research Center of the Andrew Young School of Policy Studies at Georgia State University maintains records of governmental expenditures in every school district across the state. I obtained records on nutritional expenditures broken up by local, state, and federal funding source from FY 2000 to FY 2017. Due to stipulations of tracking federal funding, these data are some of the most accurate expenditure accounts kept by the state of Georgia.[16] The data can be broken up by school and by nutrition expenditure code, which tracks the different types of nutritional outlay. However, over 95% of nutrition expenditures are on food, so I disregard the funding categories. I also disregard school-level records because these are mostly missing. Table B.29 displays nutrition expenditures from state, local, and federal sources broken up by farm-to-school policy adoption

[14] Gordanier et al. (2019). A school barely qualifying, with 40% ISP, receives reimbursement equal to 64% of expenditures. All schools with ISP above 62.5% receive full reimbursement for their breakfasts and lunches.

[15] On this public web-page.

[16] According to conversations with Nicholas Warner at the Fiscal Research Center.

status. Clearly, districts that adopt farm-to-school policies are statistically different from those that do not, being larger on average.

3.4 Student Enrollment and Demographics

Supplemental information on the total number of students across each school and school district was obtained through public records posted to the website of the Georgia Department of Education.[17] This information was linked to supplemental information sourced from the Common Core of Data using the Stata add-on Education Data Portal package developed by the Urban Studies Institute.[18] This information is summarized in Table B.29.

3.5 Agricultural Revenues

The Farm Gate Values Survey, maintained by the Center for Agribusiness Economic Development at the University of Georgia, maintains records of agricultural revenues across 160 Georgia counties from 2000-2018. The information is collected by agents within each county and accurately reflects total revenues in each county region.[19] The information includes revenues across 89 different commodity categories, with many commodity categories further broken down by type of grow technique. For this analysis I eliminate products likely unrelated to school nutrition.[20] The third panel of Table B.29 displays information on agricultural revenues broken up by type of product. Although school districts adopting farm-to-school policies differ from those that do not, agricultural revenues across both types of county are similar and not statistically distinguishable

[17]The data reporting tab on this webpage sub-links for reports on enrollment and free-and-reduced price lunch status.

[18]Education Data Portal (Version 0.3.0), Urban Institute, Center on Education Data and Policy, accessed May, 1st, 2019, https://educationdata.urban.org/documentation/US Department of Education Common Core of Data/

[19]In 2017, the FarmGate Values Survey reports slightly over $10 billion in revenues for agricultural products, while for the same year the US Census of Agriculture reports that the market value of all agricultural products sold in the state was $9.5 billion. This suggests that the Farm Gate Values survey accurately captures all agricultural products sold, and that the survey may in fact be a more accurate accounting of farm revenues than the national agricultural census.

[20]Products considered not relevant to school nutrition expenditures are excluded from the analysis in all regressions. These products include timber, camping, Christmas trees, corn mazes, crop insurance, fishing, horses, goats, flight quail, meat quail, tobacco, wildlife observation, government payments, and hunting leases for deer, duck, and turkey.

from each other in any commodity category used by this study. The products included in each class of commodity, such as fruit, vegetable, dairy, or meat, are listed in Table B.35.

4 Empirical Strategy

This paper exploits variation in the timing of farm-to-school policy adoption and student population affected to assess the role school nutrition expenditures and sourcing policies play on local agricultural revenues. I present the results of four statistical models: a naive panel fixed-effects OLS regression resembling a difference-in-differences model, a spatial lagged panel fixed effects model, and a triple difference model relying on plausibly exogenous adoption of the Community Eligibility Provision across schools. Strengths, weaknesses, and identifying assumptions of each model are discussed below.

4.1 Naive OLS Regression Model

I first present a straightforward empirical model relying on variation in the timing of farm-to-school policy adoption across school districts. Let y_{ipt} represent agricultural revenues in county i for agricultural product p in year t, where product p is often defined as all agricultural products.

$$y_{ipt} = \beta FTS_{it} + Z_{it}\gamma' + \eta_i + \tau_t + \varepsilon_{ipt}. \tag{3.1}$$

In Equation 3.1, FTS_{it} is a variable representing adoption of farm-to-school policies by school districts within a county. Because school districts do not perfectly correspond to counties, this term is a continuous variable representing the proportion of a county's public K-12 population served by farm-to-school-adopting districts in a school year. Z_{it} controls for time-varying total student population within the county and free and reduced lunch shares. η_i is a county fixed effect and τ_t is a year fixed effect. These fixed effects control for baseline differences in agricultural revenues across counties and secular changes in agricultural revenues across the entire state over time.

Identifying Assumptions: To predict the causal effect of farm-to-school policies on local agricultural revenues, Equation 3.1 requires that counties with more farm-to-school adopting districts would have had similar counter-factual post-adoption trends in agricultural revenues as counties with no (or fewer) adopting districts. An important potential violation of this assumption would be if county populations have changing preferences over time that are correlated both with agricultural revenues and the timing of farm-to-school policy adoption. For example, counties with more adopters of the policy may experience simultaneous increases in purchases from farmers markets, or retailers may change their sourcing patterns at the same time as the policy adoption. Equation 3.1 also requires that the expectation of the error term, ε_{ipt}, is zero conditional on the model covariates. This is unlikely due to spatial correlation and policy spillovers. To start with spatial correlation, agricultural revenues are inherently place-based; agglomeration effects, land suitability and availability, and distance to markets all affect the location and magnitude of agricultural revenues. The county unit of observation does not perfectly overlap with these locational factors, so we may expect regionally-correlated agricultural revenues to violate the assumptions of the model. Moreover, time-varying factors in productivity may cause regional perturbations in the error term, also violating the model's assumptions. Next, policy spillovers occur when school districts sourcing from "local" vendors source from nearby counties. This tendency would attenuate the estimated policy change related to farm-to-school adoption because counties without the policy change may experience simultaneous increases in agricultural revenues. Finally, there is evidence that local production increases may in fact cause these policy shifts, as school districts re-allocate expenditures where local supply permits.[21] This reverse causation is another important potential violation of the identification assumption.

4.2 Spatial Lag Model

To address potential issues with spatial correlation and policy spillovers raised above, I present a spatial lag model that allows for regionally-correlated agricultural revenues and regionally-

[21] O'Hara and Benson (2019).

correlated error terms. Let y_{ipt} represent agricultural revenues in county i for agricultural product p in year t, where product p is often defined as all agricultural products.

$$y_{ipt} = \lambda \sum_{j=1}^{n} w_{ij} \cdot y_{jpt} + \beta FTS_{it} + Z_{it}\gamma' + \eta_i + \tau_t + \varepsilon_{ipt}. \tag{3.2}$$

In Equation 3.2, $\sum_{j=1}^{n} w_{ij} * y_{jpt}$ is a weighting term that controls for variation in agricultural revenues in all counties in the state, y_{jpt}, where the weighting term w_{ij} is the inverse distance between county i and county j. The term λ measures the correlation in agricultural revenues across county i and counties j. As before, FTS_{it} is a variable representing adoption of farm-to-school policies by school districts within a county, Z_{it} controls for time-varying student population and free and reduced price lunch shares, η_i is a county fixed effect, and τ_t is a year fixed effect. In this model, ε_{ipt} is a spatially autoregressive error term that allows agricultural revenue perturbations to be affected by error term disturbances in nearby counties, where I define "nearby" counties $j \in \{1, ..., n\}$ as all contiguous counties.[22]

Next, I present a spatial lag model that allows direct testing of effects of nutrition expenditures and farm-to-school policy adoption on the agricultural revenues of contiguous counties. Let $\mathbf{Y}$ represent y_{ipt} and $\mathbf{W}$ represent the lagging weights $\sum_{j=1}^{n} w_{ij}$. Consider the following (equivalent) spatial regression specifications:

$$\begin{aligned} y_{ipt} = \lambda \sum_{j=1}^{n} w_{ij} \cdot y_{jpt} + \beta_1 FTS_{it} + \beta_2 \sum_{j=1}^{n} w_{ij} \cdot FTS_{jt} \\ + Z_{it}\gamma_1' + \sum_{j=1}^{n} w_{ij} \cdot Z_{jt}\gamma_2' + \eta_i + \tau_t + \varepsilon_{ipt} \end{aligned} \tag{3.3}$$

[22]Note that the error term disturbances are allowed to be correlated only with contiguous counties, while the dependent variable is lagged across all counties in inverse proportion to distance. I adopt different contiguity matrices because the year controls handle error term disturbances that are constant across all counties but do not control for sub-regional fluctuations (for example, in the case of a drought). The contiguous-county error matrix allows for such correlations. Spatially correlated agricultural revenues, meanwhile, are influenced by markets in proportion to their distance to those markets and not simply by the average revenues in contiguous counties, so it seems more reasonable to adopt an inverse-distance contiguity matrix.

$$\mathbf{Y} = \lambda\mathbf{WY} + \beta\mathbf{WFTS} + \gamma\mathbf{WZ} + \varepsilon \tag{3.4}$$

In Equation 3.3, the weighting term $\sum_{j=1}^{n} w_{ij} * y_{jpt}$ is the same as before. Likewise, FTS_{it} is a variable representing adoption of farm-to-school policies by school districts within a county, Z_{it} controls for time-varying county characteristics including total student population and free and reduced price lunch shares, η_i is a county fixed effect, τ_t is a year fixed effect, and ε_{ipt} is a spatially autoregressive error term that allows agricultural revenue perturbations to be affected by error term disturbances in nearby counties. The two new terms, $\sum_{j=1}^{n} \cdot FTS_{jt}$ and $\sum_{j=1}^{n} \cdot Z_{jt}$, allow testing for policy spillover effects in nearby counties. β_2 measures the extent to which farm-to-school policy adoption affects agricultural revenues in nearby counties, while γ_2 is the relationship between each time-varying covariate in the vector Z_{jt} and agricultural revenues. Of special interest, γ allows testing the extent to which local nutrition expenditures impact agricultural revenues of contiguous counties.

Identifying Assumptions: To predict the causal effect of farm-to-schools policies on local agricultural revenues, Equation 3.3 requires that counties with more nearby farm-to-school adopting districts would have had similar counter-factual post-adoption trends in agricultural revenues as counties with no (or fewer) nearby adopting districts. As before, one violation of this assumption is the selection bias associated with adopting districts having unrelated increases in agricultural revenues correlated with the timing of farm-to-school policy adoption. Moreover, the possibility of reverse causation remains. To assess the importance of selection bias and reverse causation, I next turn to an empirical model relying on exogenous changes in another school nutrition policy, the community eligibility provision of the Healthy Hunger Free Kids Act.

4.3 Community Eligibility Provision

To secure plausibly causal identification, I employ a related empirical model that exploits simultaneous variation in two policies. The first is school-district adoption of farm-to-school policies. The second is the Community Eligibility Provision (CEP) of the Healthy Hungry Free Kids Act (HH-

FKA), which allows schools with over 40% of their student population qualifying for some form of federal aid to serve free lunch to all their students, saving administrative burden. Although many schools are eligible CEP schools, only some actually participate. Participating schools experience an increase in the number of meals that they serve to students as marginal students are induced into eating cafeteria food. These increased nutritional outlays are primarily financed by the federal government, making it unlikely that coincidental local tax changes may indirectly affect agricultural revenues. Plausibly causal identification comes from the fact that, when faced with the need and ability to purchase more school food, farm-to-school districts may be more likely to source it locally. Figure A.19 shows the increase in expenditures among districts with CEP-participating schools. Table B.37 empirically demonstrates that CEP adoption reliably increases nutrition expenditures across every definition of CEP participation.

Let y_{ipt} represent agricultural revenues in county i for agricultural product p in year t, where product p is often defined as all agricultural products.

$$y_{ipt} = \beta FTS_{it} \cdot CEP_{it} + Z_{it}\gamma' + \eta_i + \tau_t + \varepsilon_{ipt} \tag{3.5}$$

Terms FTS_{it}, Z_{it}, η_i, τ_t, and ε_{ipt} are defined as before. The term CEP_{it} represents the share of students marginally induced into free meals according to school-level adoption of the Community Eligibility Provision.[23] Rather than identifying CEP_{it} off potentially endogenous participation in the program, the term represents the county share of students induced into "free meal" status by adoption of the program. For example, a county in which every student qualifies for free-and-reduced price lunch before the HHFKA, and in which all schools adopt the CEP, would receive a value of 0 because no student is marginally induced into free-meal status. Likewise, a school district in which no student elects to receive free-and-reduced price meals before the HHFKA, where all schools within the county adopt the CEP, would have a CEP_{it} value of one. In practice, neither situation occurs.

[23]For each county, $\text{CEP}_{it} \equiv \sum_{s=1}^{n}(1 - FRL_{st}) \cdot Participation_{st} \cdot \frac{SchoolEnrollment_{st}}{CountyEnrollment_{it}}$ over all schools in a district $s \in \{1,...,S\}$ where $Participation_{st}$ is a binary variable representing school s participation in the CEP program in year t.

One might also consider a spatially-lagged version of the same empirical strategy. Let y_{ipt} represent agricultural revenues in county i for agricultural product p in year t, where product p is often defined as all agricultural products.

$$
\begin{aligned}
y_{it} = \beta_1 FTS_{it} + \beta_2 CEP_{it} + \beta_3 FTS_{it} \cdot CEP_{it} + \lambda_1 x_{it} + \lambda_2 \sum_{j=1}^{n} w_{ij} \cdot y_{jt} \\
+ \psi_1 \sum_{j=1}^{n} w_{ij} \cdot FTS_{jt} + \psi_2 \sum_{j=1}^{n} w_{ij} \cdot CEP_{jt} + \psi_3 \sum_{j=1}^{n} w_{ij} \cdot FTS_{jt} \cdot CEP_{jt} \\
+ Z_{it} \gamma_1' + \eta_i + \tau_t + \varepsilon_{it}
\end{aligned}
$$

Terms FTS_{it}, CEP_{it}, Z_{it}, η_i, τ_t, and ε_{ipt} are defined as before. This regression model synthesizes the identification strategies presented in Equation 3.3 and Equation 3.5, allowing potentially unbiased estimation of the effect of nutrition expenditures on local agricultural revenues and the effect of farm-to-school policy adoption on local agricultural revenues, where local includes both the same county and contiguous counties. A pitfall of this well-identified model is that CEP variation is only present for a short three-year period out of the 17 year sample.

Identifying Assumptions: To estimate the causal relationship between nutrition expenditures and farm-to-school policy adoption and local agricultural revenues, Equation 5 and Equation 6 require that there are no county-specific factors correlated both with the share of students induced into "free" meal status and farm-to-school policy adoption across districts. Such correlated factors must also impact local agricultural revenues. In Equation 6, the correlated factors must be county-specific, as local revenues are allowed to fluctuate with regional agricultural revenues and regional disturbances in agricultural revenues. It seems unlikely that some factor may be correlated with both of these variables simultaneously and not be controlled by county fixed effects, time-varying county characteristics, and regional lags and correlated errors. The primary shortcoming of this identification model, rather, is that a relatively small share of students are induced into free meal status over a relatively short time frame. This weakness reduces the likelihood of observing statistically significant relationships.

4.4 Two Stage Least Squares – Share of CEP Students

The empirical strategies laid out above may still be susceptible to bias from reverse causation. To eliminate this bias, I employ a two stage least squares regression relying on variation in the share of students induced into free and reduced price lunch status. Farm-to-school adoption may be endogeneous because districts select into the policies, so I do not include any interaction with farm-to-school policy adoption. Let y_{ipt} represent agricultural revenues in county i for agricultural product p in year t, where product p is often defined as all agricultural products. Let x_{it} represent total school district nutritional expenditures in county i and year t.

$$x_{it} = \beta CEP_{it} + Trend_i'\delta + \eta_i + \tau_t + u_{it}$$
$$y_{it} = \gamma x_{it} + Trend_i'\delta + \eta_i + \tau_t + \varepsilon_{it}$$

In the above two stage least squares model, terms CEP_{it} and η_i are defined as before. τ_t is a year dummy. $Trend_i$ is a district-specific trend control. Rather than identifying CEP_{it} off potentially endogenous student population or district decisions to participate in the program, the term represents the county share of students induced into "free meal" status by adoption of the program.[24] Intuitively, this empirical model uses variation in the share of students induced into free meal status to predict the change in nutrition expenditures, and this change in nutrition expenditures is then used to predict changes in within-county agricultural revenues. The two-stage least squares provides an estimate of the baseline relationship between agricultural revenues and nutrition expenditures with which to compare previous estimates.

Identifying Assumptions: A causal interpretation requires that the share of students induced into free lunch status strongly predicts nutrition expenditures. As before, Figure A.19 shows the increase in expenditures among districts with CEP-participating schools. Table B.37 demonstrates that CEP adoption reliably increases nutrition expenditures across every definition of CEP participation. Moreover, the share of students induced into free lunch status must affect local agricultural

[24] The formula for this variable is: $CEP_{it} = \frac{\sum_s CEP_{st}(1-FRL_{st})}{Enrollment_{it}}$.

revenues only through the changes in nutrition expenditures, otherwise known as the exclusion restriction. Although the exclusion restriction cannot be directly tested, it seems unlikely that small variations in the free-and-reduced lunch population across CEP-participating schools would directly affect local agricultural revenues except through school nutrition expenditures.

5 Results

Table B.30 presents results across all empirical specifications from section 4, including five regression models testing for local agricultural revenue increases from baseline school district nutrition expenditures and five regression models testing for local agricultural revenue increases related to adoption of farm-to-school policies. Across all specifications, I find weak statistical evidence that school nutrition expenditures are associated with local agricultural revenues. For each dollar spent on school nutrition, agricultural revenues in the same county are \$0.13 to \$0.19 higher. This relationship is not necessarily causal; since the Farm to School Census found that 17% of district expenditures in farm to school districts goes to local farmers, these point estimates are not unreasonable but are likely biased by selection and possible reverse causation. Additionally, these findings should be relatively higher because the spending multiplier on local food is 1.5 according to Christensen et al. (2018).

Columns (1) and (2) of Table B.30 present the results of a naive panel fixed effects model, where in column (2) the variable representing farm-to-school policies is an interaction between farm-to-school adoption and post-adoption year weighted to the county by student population. The coefficient on farm-to-school policy adoption, "FtS Policy," is relatively large but not statistically significant. The point estimate 1913, which is expressed in thousands, suggests that each post-policy adoption year is associated with increases in local agricultural revenues of \$1.9M. Scaling this average across all policy years implies agricultural revenue increases of \$550M related to farm-to-school policies.[25] Although this is small in comparison to the \$167B in agricultural revenues

[25]The average effect over 68 counties with an average of 4.2 post-policy years, or 288 total post-policy district years, can be converted to the total implied effect by multiplying the point estimate times 288 district years and then 1000

from 2001-2017, it represents about 3.9% of all nutrition expenditures over the sample, \$13.8B.

Columns (3) through (6) of Table B.30 display results of empirical specifications discussed in section 4.2. In these models, variations in county agricultural revenues are lagged by revenues in all other counties in the state in proportion to the inverse distance between county i and j. Moreover, error term disturbances in contiguous counties are allowed to be correlated. Columns (3) and (5) show the baseline relationship between nutrition expenditures in county i and agricultural revenues in both county i and in counties j that are contiguous to county i. The point estimates suggest that roughly \$0.15-\$0.18 of every dollar spent on school nutrition is recouped by farmers in the same county, while a surprisingly negative \$0.94 is lost by farmers in bordering counties. It is unclear why local expenditures are associated with negative revenues in bordering counties, although this may relate to unobservable agricultural differences in urban and suburban regions with greater nutrition expenditures. Columns (4) and (6) jointly estimate the effect of nutrition expenditures and farm-to-school policies on agricultural revenues in local and nearby counties. Although these models do not control for selection into the farm-to-school policies, they accurately control for regionally correlated revenue levels and error-term disturbances. Controlling for these error disturbances dramatically improves the precision of the estimated relationship between farm-to-school policy adoption and local agricultural revenues; these point estimates are similar to those of column (2), although they're not quite statistically significant. The farm-to-school policy variable is consistent across both models, suggesting large local economic effects of slightly over \$550M. The coefficient on "FtS Policy" in the sixth row suggests that perhaps \$6.5M in nutrition spending was lost by neighboring counties, although this term is not significant.

The final four columns of Table B.30 display results of the empirical specifications discussed in section 4.3. These estimates rely on variation in nutrition spending resulting from a federal policy change that is unrelated to farm-to-school adoption, plausibly purging the selection bias associated with school districts adopting farm-to-school programs or reverse causation. The variable CEP is the share of a county's students that are induced into free lunch status by the new policy.[26] Clearly,

because the point estimates are expressed in thousands.

[26]As shown in Table B.37, school-level adoption of the CEP policy in general is strongly statistically associated

increasing shares of students induced into free meal status is strongly and statistically significantly associated with increases in local agricultural revenues. The coefficient of 43,000 in column (8), for example, means that going from no students induced into free lunch to all students induced into free lunch would increase local revenues by $43M. Of course, it is impossible for the CEP variable to increase by more than 0.4, and on average this variable is in fact 0.005. Therefore, on average, the share of marginally induced free lunch students was associated with a $215,000 increase in local agricultural revenues. Scaling by all marginally induced free lunch CEP students, this is a total change of 3.3M. In contrast, the smaller point estimate of 3,365 in column (8) suggests that each year of farm-to-school adoption is associated with $3.36M increase in local agricultural revenues. Over all farm to school years, this represents $967M in increased local revenues. This value should be corrected by the point estimate of $CEP*FtS$, -76,000. Since the average value of $CEP*FtS$ is 0.001 and the total of all such interactions is 3.8, the figure of $967M should be scaled back to $678M.[27] That is, the linear combination of the terms in row (2) and row (4) in columns (8) and (10) suggests that farm to school programs are associated with agricultural revenue increases of roughly $680M.

Surprisingly, results in the second panel of Table B.30 provide little clarity on the effect of agricultural spending on contiguous counties. The only statistically significant coefficients are the effects of increasing CEP shares, which are positive. These would suggest that increased nutrition expenditures from CEP adoption are associated with large increases to agricultural revenues in contiguous counties. The coefficient of 85,222 in column (9), for example, suggests that going from zero students induced into free lunch to all students induced into free lunch would increase revenues in contiguous counties by $85M. Since the typical district's marginal share of CEP students is 0.005, this implies that a typical district's increase in nutrition expenditures associated with CEP adoption may have led to an average increase in agricultural revenues in contiguous counties of $425,000. The total increase in revenues in contiguous counties would therefore be $323M. The negative coefficient on the farm-to-schools variable in the second panel, meanwhile, may suggest

with increases in nutrition expenditures.

[27] $678.2 = 967 - 3.8 * 76M$

that districts are substituting to more-local suppliers from more-distant providers, although this coefficient is not statistically distinguishable from zero.

I present a subset of regression models across animal products, fruits and vegetables, and agrotourism in Table B.31, Table B.32, and Table B.33. Animal products, which includes revenues for beef, catfish, chicken, dairy, eggs, fishing, and pork, appear related to baseline nutrition expenditures, although the exact magnitude of the coefficient fluctuates between an insignificant point estimate of \$0.08 per dollar spent to the \$0.229 observed in column (5). Farm-to-school policies appear strongly associated with increases in local revenues for animal products, although it is unclear how much of the significant effects in columns (3) through (6) may be attributed to selection and reverse causation. It seems plausible that a school district with nearby milk or egg production would be more inclined to enact farm-to-school policies. The fact that coefficients on the CEP share, in columns (5) and (6), are not significant may provide support for the notion that the farm-to-school variable is more endogenous and prone to reverse causation for animal products.

The results of Table B.32 provide weaker evidence that nutrition expenditures increase local agricultural revenues on fruits and vegetables, with perhaps \$0.01 to \$0.14 cents of each dollar spent by school districts recouped by local farmers. The coefficients on the farm-to-school policy variables, however, are universally negative, often statistically significantly so. Unlike the results for animal products, however, increasing CEP shares are associated with increases to local agricultural revenues. It is unclear how to interpret these results. Table B.33 shows how revenues for agrotourism are associated with farm-to-school policy adoption. Since school field trips to farms are a major element of farm-to-school policies, with a typical district engaging in 13 separate such field trips in any given year, it seems likely that agrotourism revenues might be associated with policy adoption. However, each point estimate on the farm-to-schools variable in the table is negative and statistically insignificant. Although the interaction $CEP * FtS$ is significant and positive, it is unclear why this interaction should be positive because CEP shares are not related to field trip visits a priori.

Finally, to provide an estimate of the baseline relationship between nutrition expenditures and agricultural revenues, Table B.34 depicts the results of a two stage least squares estimation procedure. Clearly, increasing shares of students receiving free lunch through the CEP is associated with dramatic increases to school nutrition expenditures. The first-stage coefficient of 2.6M suggest that a typical district's CEP share increased local nutrition expenditures by $13,000 in each year. Moreover, I find evidence that these increasing expenditures were in part recouped by local farmers. The coefficient of 26.37 in column (1) suggests that a typical county's change to nutrition expenditures associated with the CEP increased local agricultural revenues by $350,000. This value is roughly 7% of the average district expenditures in each county. If we extrapolate this value over the entire sample window, it suggests that $966M of school district expenditures flowed to local within-county producers. This is 11.3% of all school district expenditures; interestingly, this figure is less than the local share found in the 2015 Farm to School Census, which was 17%.[28] Since my estimates relate to a local variable of only within the same county, this makes sense. The estimated relationship is strongest for fruits and vegetables, not animal products; this may suggest that within-county expenditure shifts are more likely drivers of fruit and vegetable revenues than animal product revenues, as economical large-scale feeding operations are less likely to be present locally.

5.1 Discussion of Implied Revenue Changes

The Farm to School Census reports that $40M was spent on local food by Georgia school districts in 2015.[29] Extrapolating this value over the 17 year sample would be $680M, although it is unlikely that this level of local investment was constant over the entire sample period. The point estimate on "FtS Policy" in the naive regression model, reported in column (2) of Table B.30, would suggest that this figure is the reduced $550M. Since this estimated relationship does not account for spatial correlation, spatial lags, selection bias, or reverse causation, it seems reasonable that the estimate is noisy, although it is surprising that is so close to the surveyed quantity. In columns (4) and (6) of

[28]USDA (2017).
[29]USDA (2017).

Table B.30, the point estimate on “FtS Policy” implies an increase of \$555M and \$627M in local revenues associated with farm-to-school policies. This figure is, again, smaller than what might be expected from the Farm to School Census, perhaps reflecting selection bias, reverse causation, or merely the local share of spending being lower in earlier years of the sample. When incorporating variation from the Community Eligibility Provision, I find implied local revenue shifts, that are very similar to the extrapolated figure from the Farm to School Census. These figures range from \$627M to \$683M, which are almost identical to \$680M. It should be the case, however, that these figures are higher than the extrapolated value in the farm-to-school census, as the census only received replies from 84% of Georgia school districts.[30] Despite this fact, it would make sense if these values are roughly similar if the share of local expenditures has increased over time and the 2015 surveyed value is not representative of all years from 2001-2017.

The annual revenues for all agricultural products has hovered around \$10B throughout the 2010s, up \$1.8B from the (inflation-adjusted) annual revenues in 2001. The total agricultural revenues generated from 2001-2017 is \$167B. Meanwhile, the annual school nutrition expenditures over the same period have increased from \$650M to roughly \$900M, totaling \$13.8B over the entire sample. If the implied revenue shifts associated with farm-to-school policy adoption are credible, then perhaps as much as 4.9% of nutritional expenditures in the state may be attributed to farm-to-school policy adoption. This figure is roughly 0.4% of all agricultural revenues in the state over the same period. Meanwhile, the estimates of Table B.34 suggest that perhaps 7% of nutritional expenditures remain in the same county, or as much as 0.6% all agricultural revenues in the state. Although these estimations may be a small share of all agricultural revenues, they are a relatively large share of what is likely spent locally by school districts (5-7%). If we assume that the two stage least squares results in Table B.34 are accurate, then farm-to-school policy adoption is actually responsible for 70% of all nutrition expenditures that remain in the same county. Interestingly, both the estimates reported in Table B.30 and Table B.34 are lower than the share of local expenditures reported in the Farm to School Census of 2015, which found that 17% of nutrition

[30]https://farmtoschoolcensus.fns.usda.gov/find-your-school-district/georgia

expenditures were spent locally in Georgia in 2015.[31] This divergence, however, makes perfect sense. The Farm to School Census sample includes only 84% of school districts, likely excluding many that would do not engage in these policies. It also incorporates local expenditures that flow to nearby counties, while my estimation procedure focuses on within-county changes.

6 Conclusion

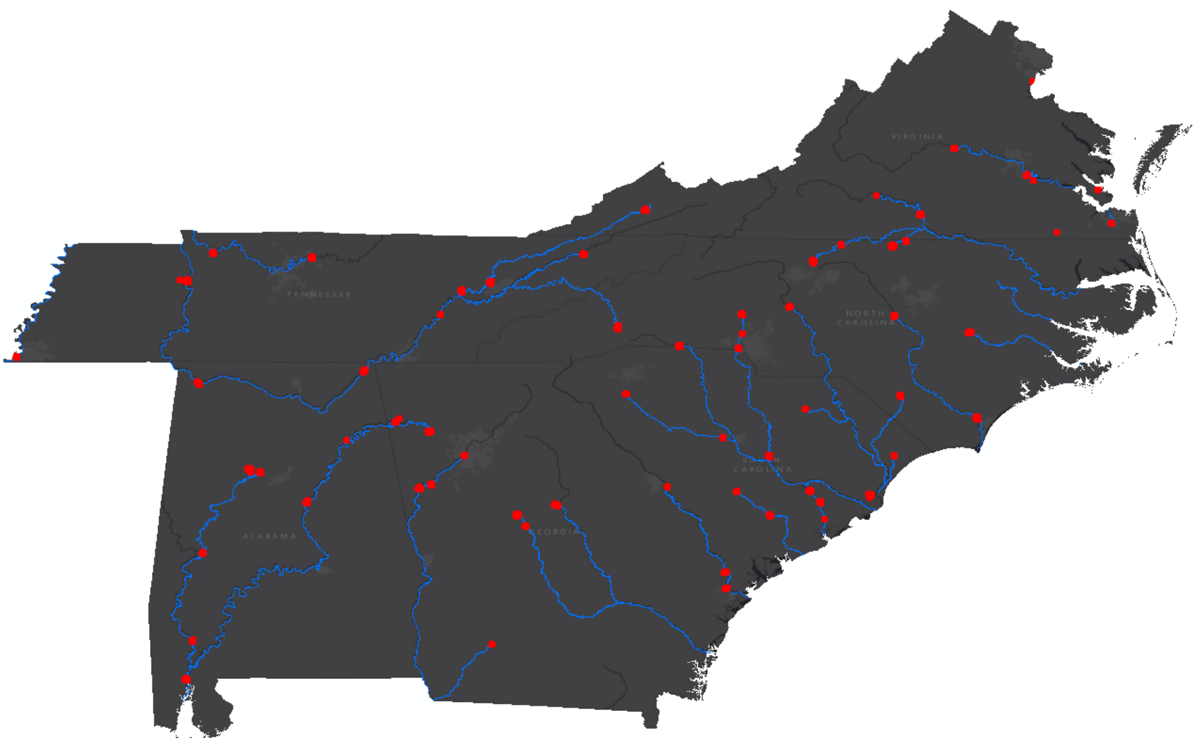
This paper provides evidence that school nutrition expenditures play an important role in local agricultural markets, whether these nutrition expenditures are associated with farm-to-school policies or not. I estimate that 5 to 7% of school district nutrition expenditures flow to within-county producers. These expenditures account for roughly 0.6% of all agricultural revenues in the state from 2001-2017, or $966M out of $167B. Of this total, perhaps as much as $680M, or 70% of the total amount of local expenditures, may be specifically attributable to farm-to-school policy adoption. My estimates for the share of local expenditures correspond almost precisely to the shares reported in the Farm to School Census after extrapolating over a longer sample. According to my estimates, roughly three quarters of the local share of school nutrition expenditures is spent on fruits and vegetables, while a smaller share of the remainder is spent on animal products. The findings suggest that school nutrition expenditures are economically meaningful drivers of agricultural markets, and local sourcing policies may be a valuable tool for assisting local farmers.

[31] USDA (2017).

Appendix A

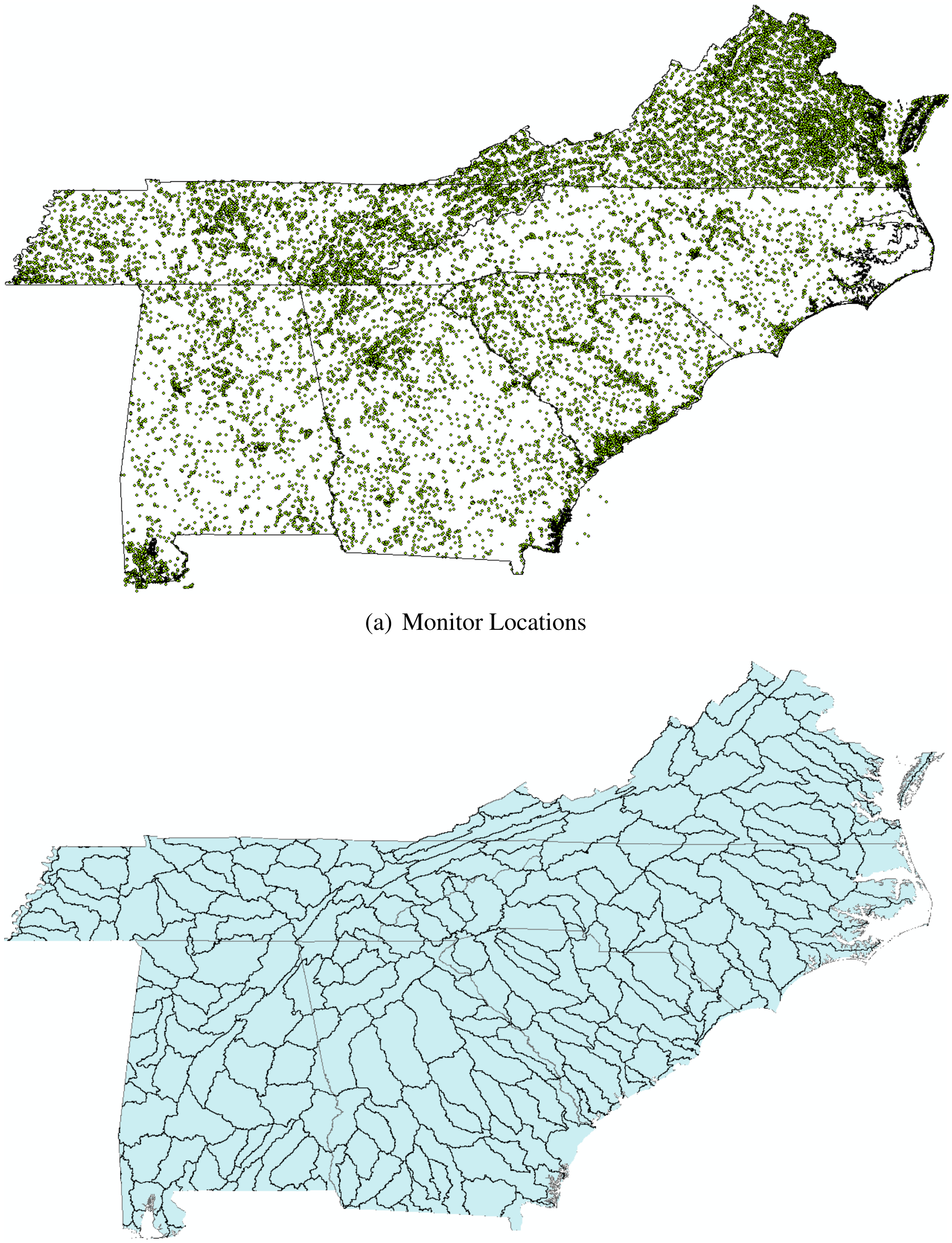
Figures

Figure A.1: Coal Ash Release Sites and Downstream River and Stream Segments



Notes: Red dots represent steam-generating coal power plants releasing a non-negative quantity of coal ash to surface waters from 2005-2017. Blue lines represent river and stream segments that are downstream from a coal ash release site.

Figure A.2: Surface Water Quality Monitoring Sites and Watershed (HUC-8) Regions

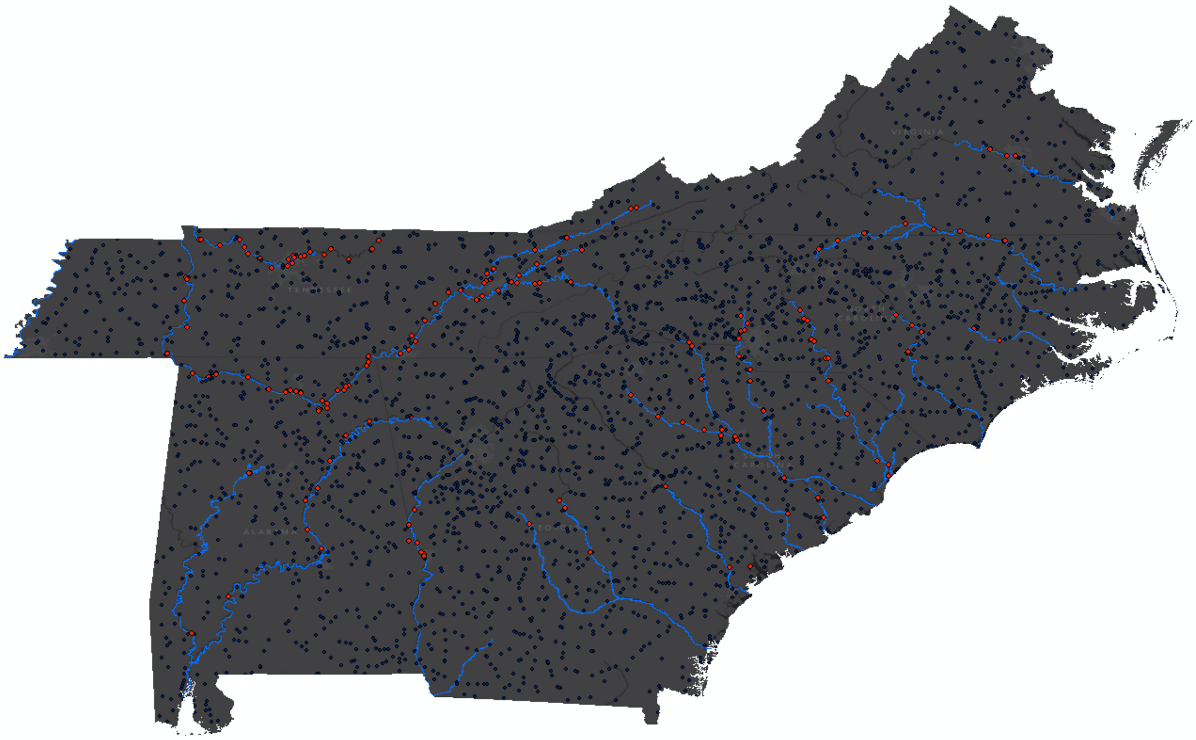


(a) Monitor Locations

(b) Watershed Regions

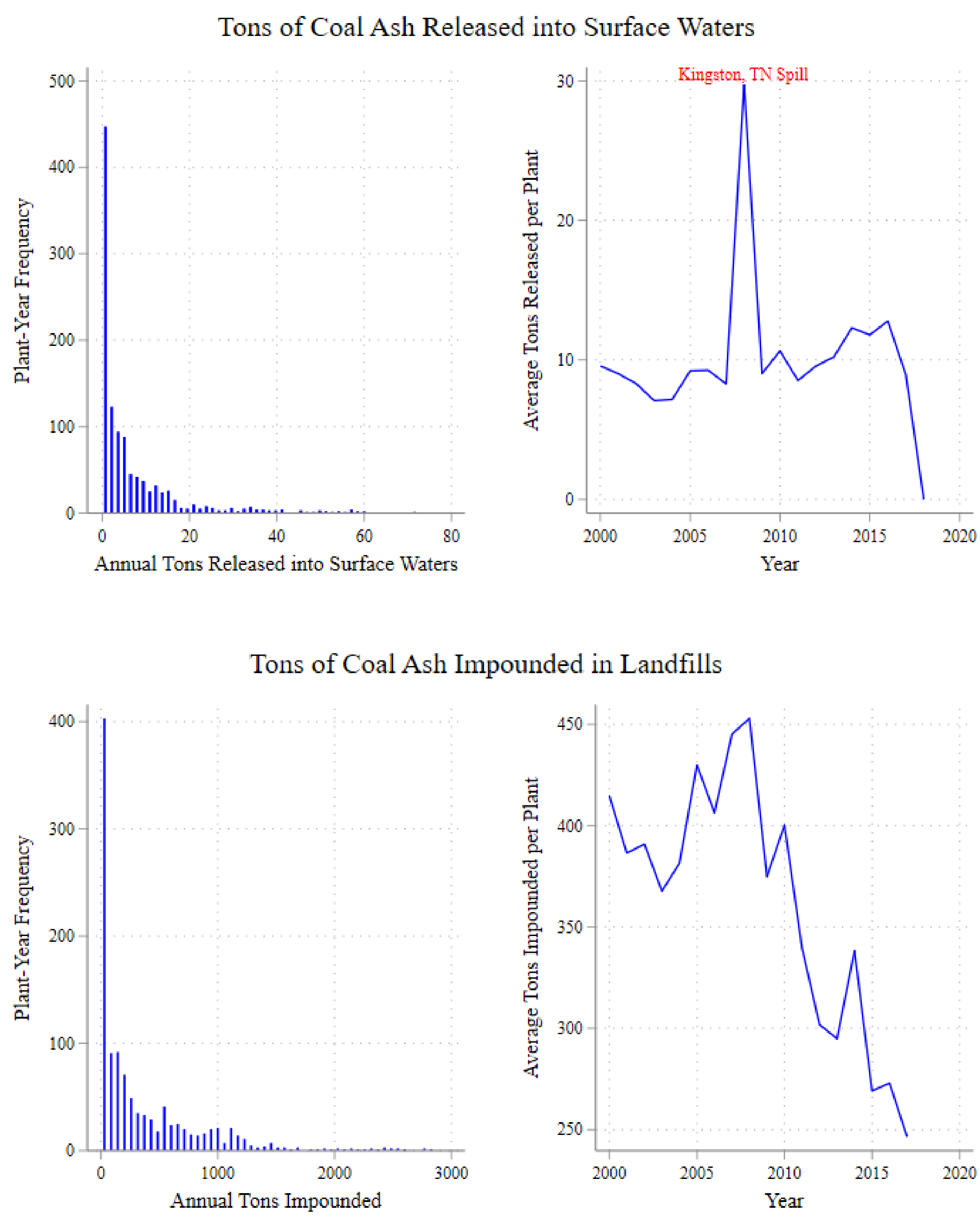
Notes: In Panel (a), green dots represent surface water quality monitor locations in the Water Quality Portal, while in Panel (b) each polygon represents a watershed of size Hydrologic Unit Code – 8.

Figure A.3: Municipal Water System Intake Locations Affected by Coal Ash



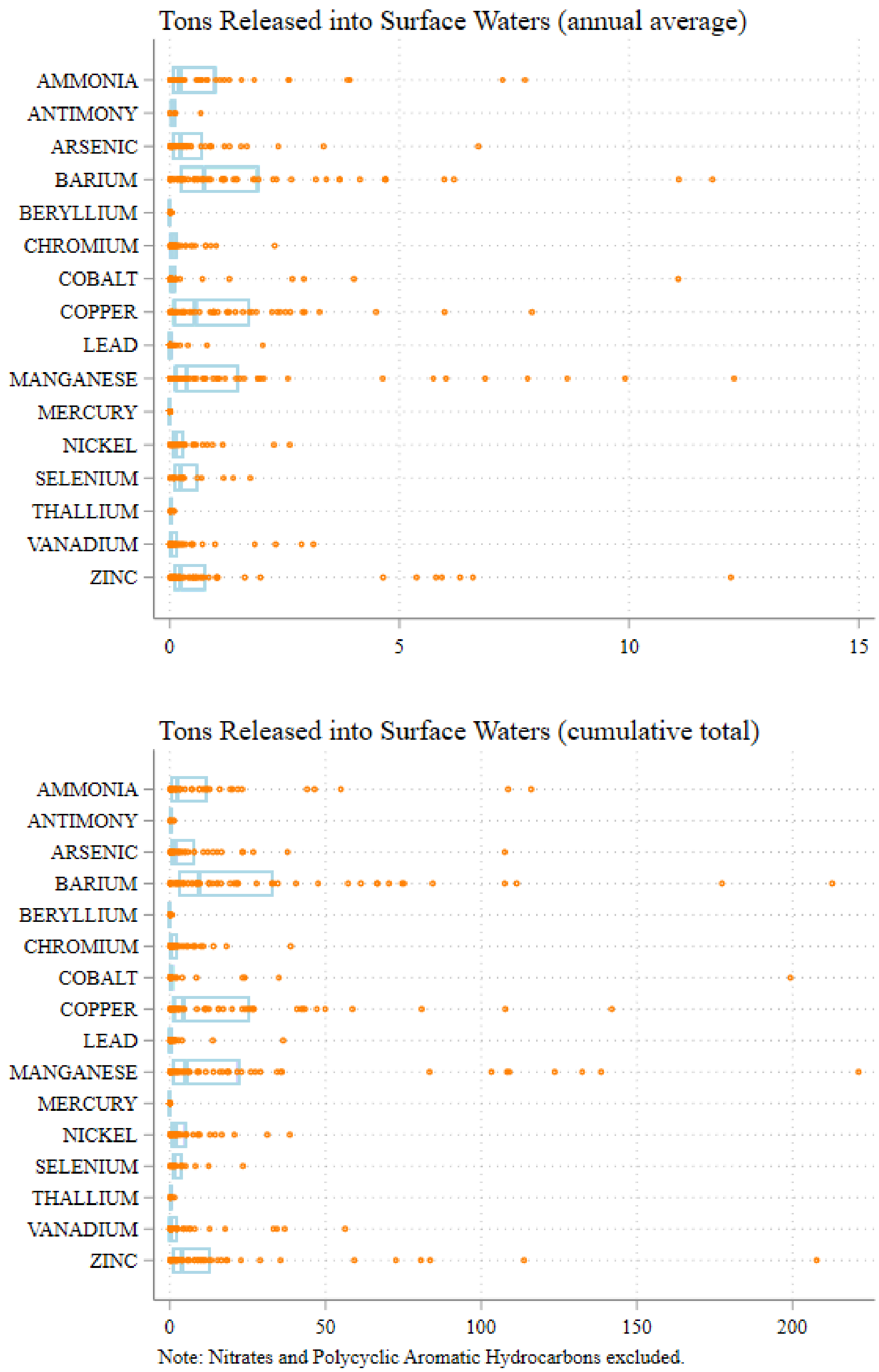
Notes: Darker blue dots represent municipal water system intake locations that are not affected by coal ash, whereas red dots are intake locations of likely affected municipal water systems according to the Southern Environmental Law Center. Blue lines represent river and stream segments that are downstream from a coal ash release site. Surface water intake locations provided courtesy of the Southern Environmental Law Center and also compiled by author.

Figure A.4: Toxic Releases by Coal Ash Plants (2000-2017)



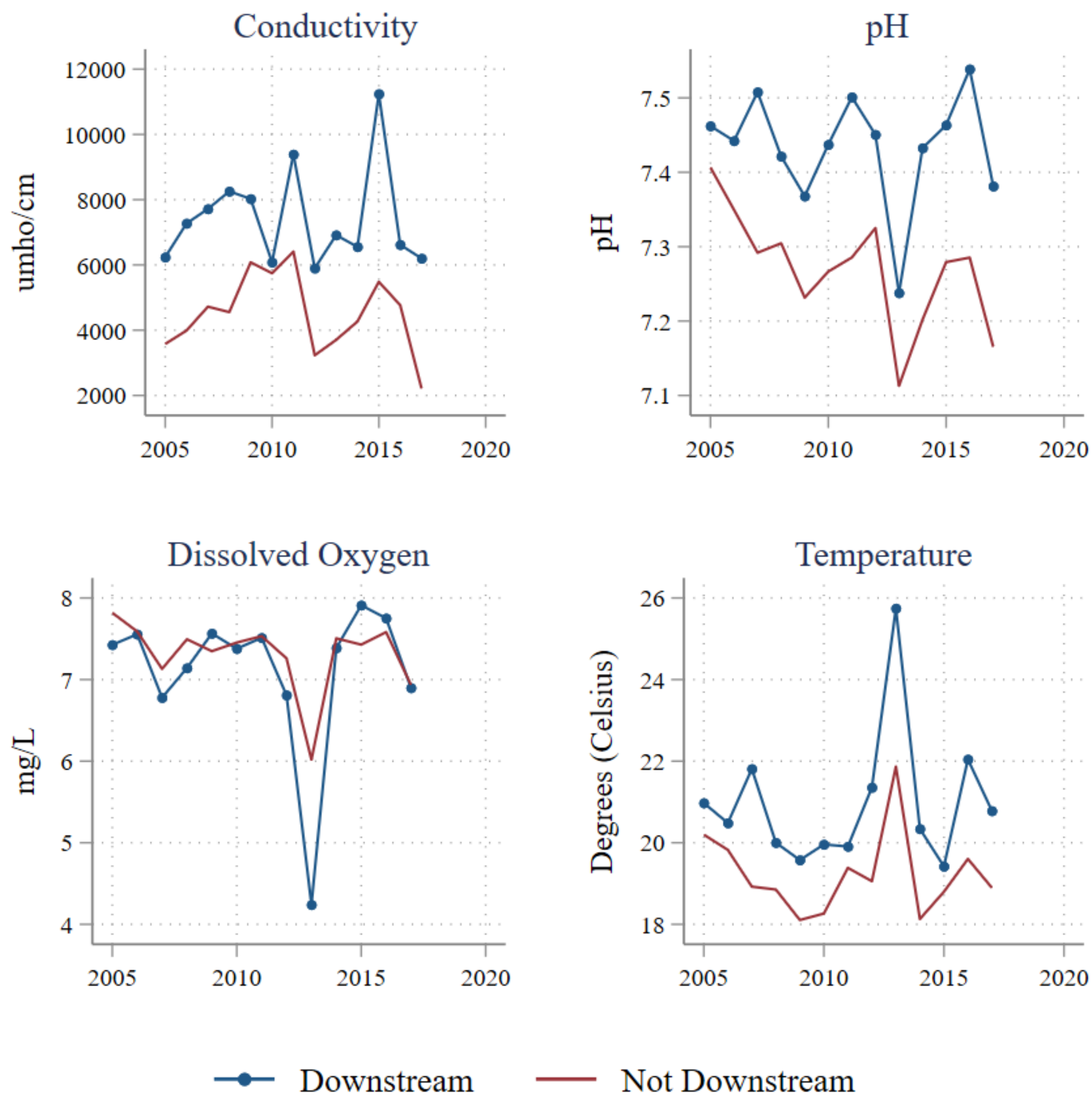
Notes: Bar charts on the left display variation in the quantity of coal ash released or impounded across all coal plants in the sample. Line charts plot the change in the quantity of coal ash effluent released into surface waters or impounded over time. Release values of zero are included.

Figure A.5: Toxic Releases by Coal Ash Plants into Surface Waters by Compound (2000-2017)



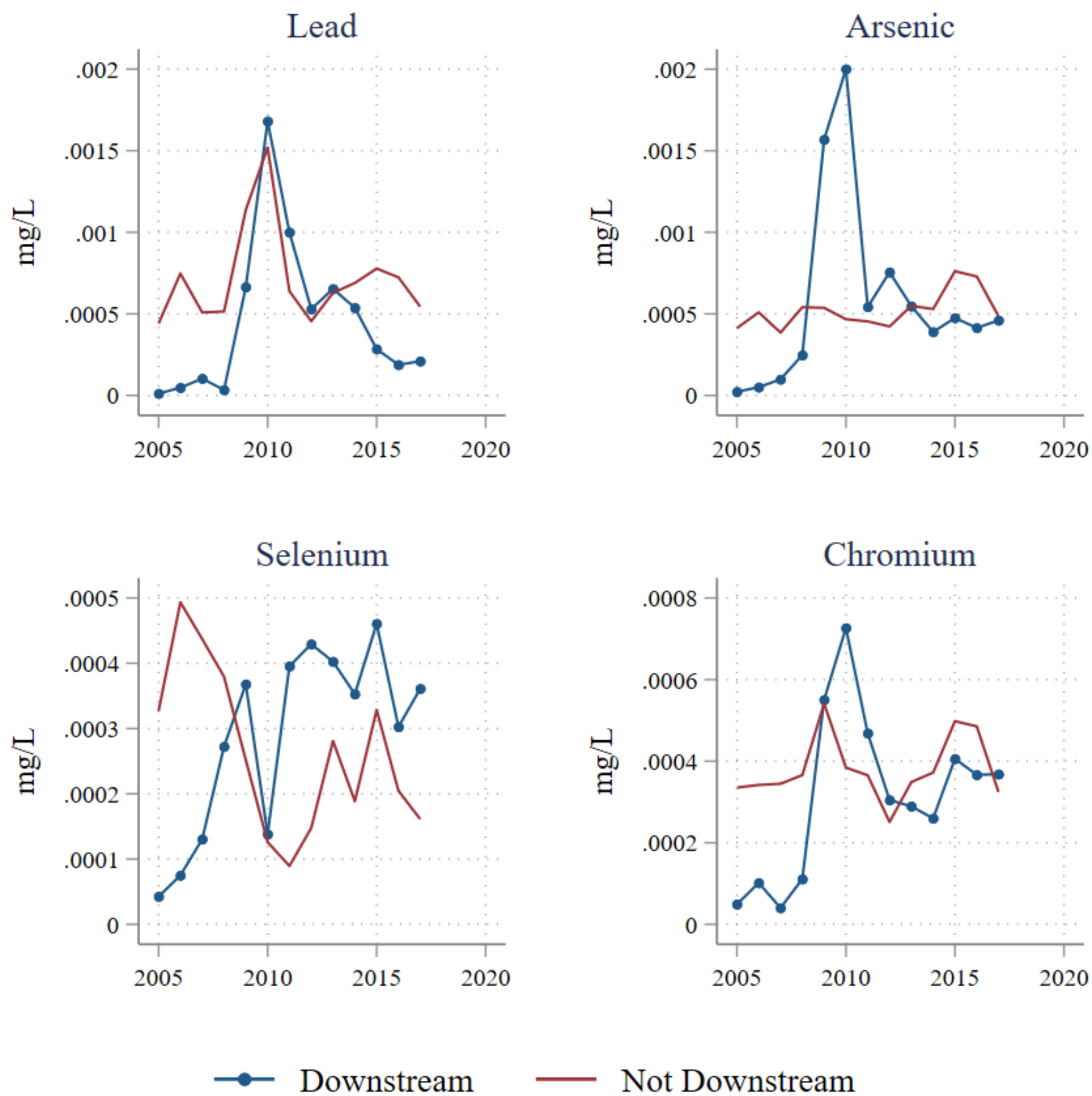
Notes: Light blue bars represent confidence intervals of the level released of each chemical across all plants, while each dot represents an individual plant observation. Release values of zero are included. Nitrates and polycyclic aromatic hydrocarbons excluded because relatively few plants release these compounds. Outliers of greater than 20 tons on average per year or greater than cumulative 300 tons are excluded for ease of visualization.

Figure A.6: Water Quality Criteria in Coal Ash Affected Surface Waters (2005-2018)



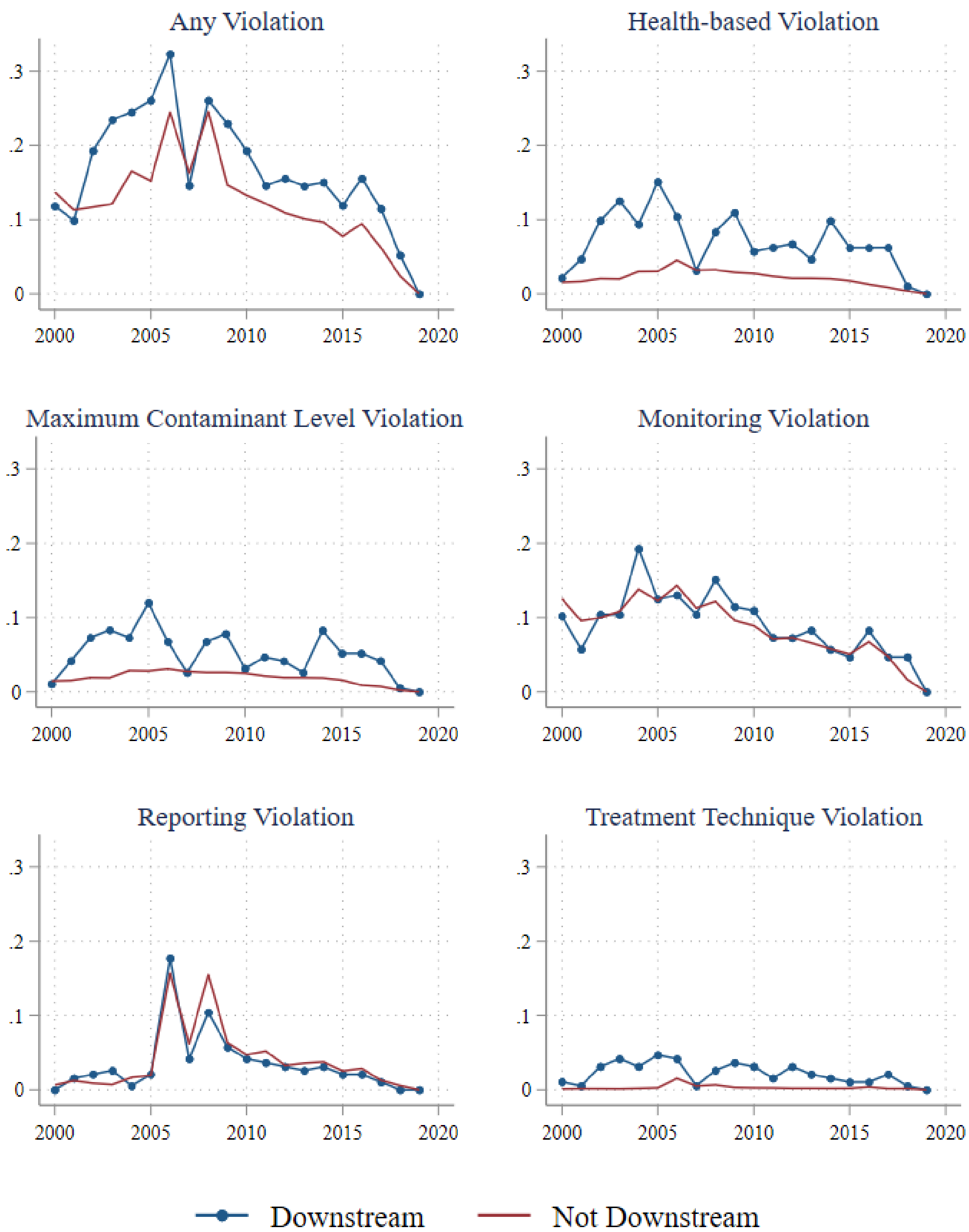
Notes: Average levels plotted across all surface water monitor tests, excluding tests of sediment and hyporheic zone. Outlier observations above the 99th percentile are excluded. The "Downstream" category includes surface water quality monitors within 25 miles downstream of a coal ash site. "Not Downstream" includes all other surface water monitors in the sample states from 2005-2017.

Figure A.7: The Concentration of Water Pollutants in Coal Ash Affected Surface Waters (2005-2018)



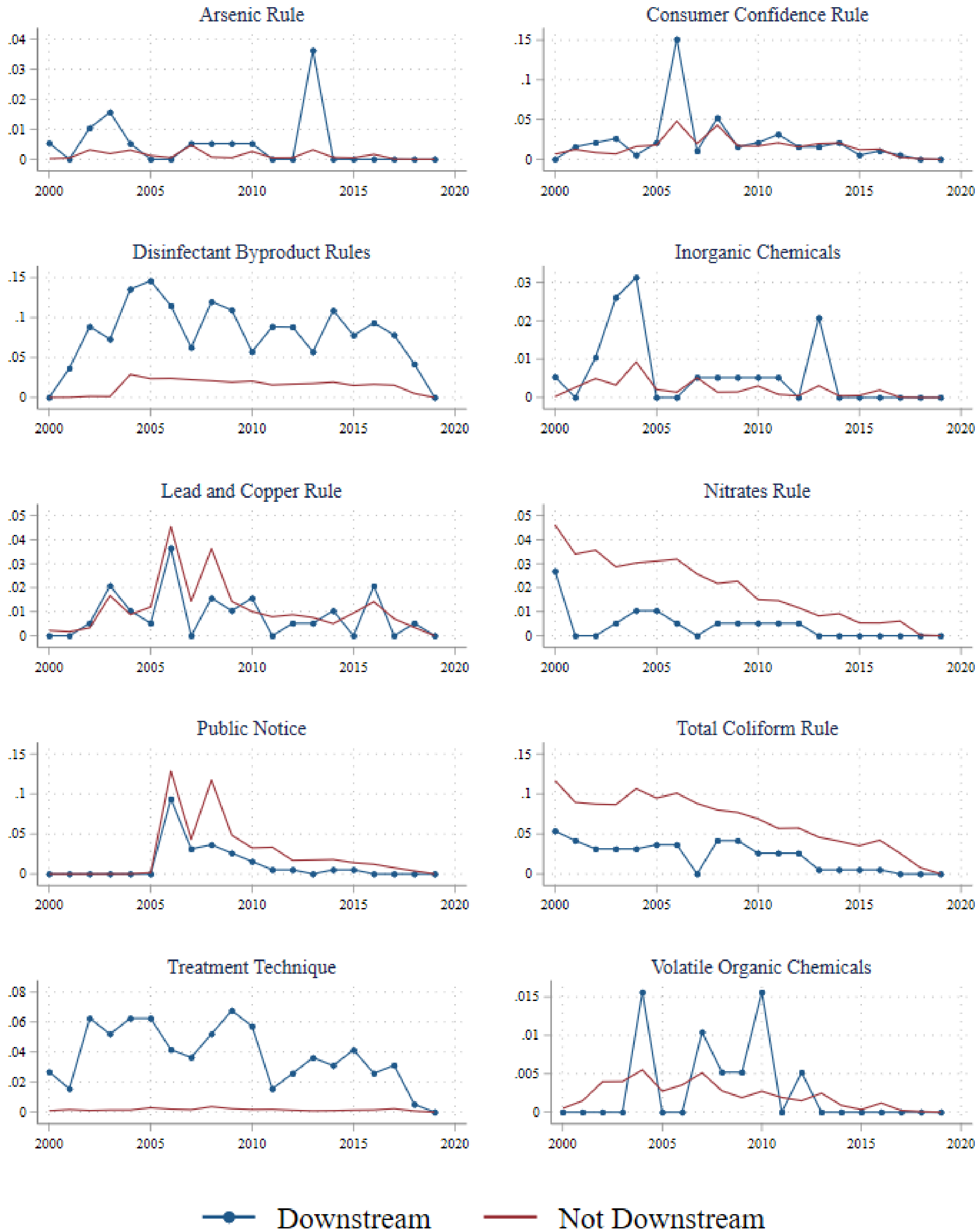
Notes: Average levels plotted across all surface water monitor tests, excluding tests of sediment and hyporheic zone. Outlier observations above the 99th percentile are excluded. The "Downstream" category includes surface water quality monitors within 25 miles downstream of a coal ash site. "Not Downstream" includes all other surface water monitors in the sample states from 2005-2017.

Figure A.8: Safe Drinking Water Act Violations by Type of Infraction (2000-2018)



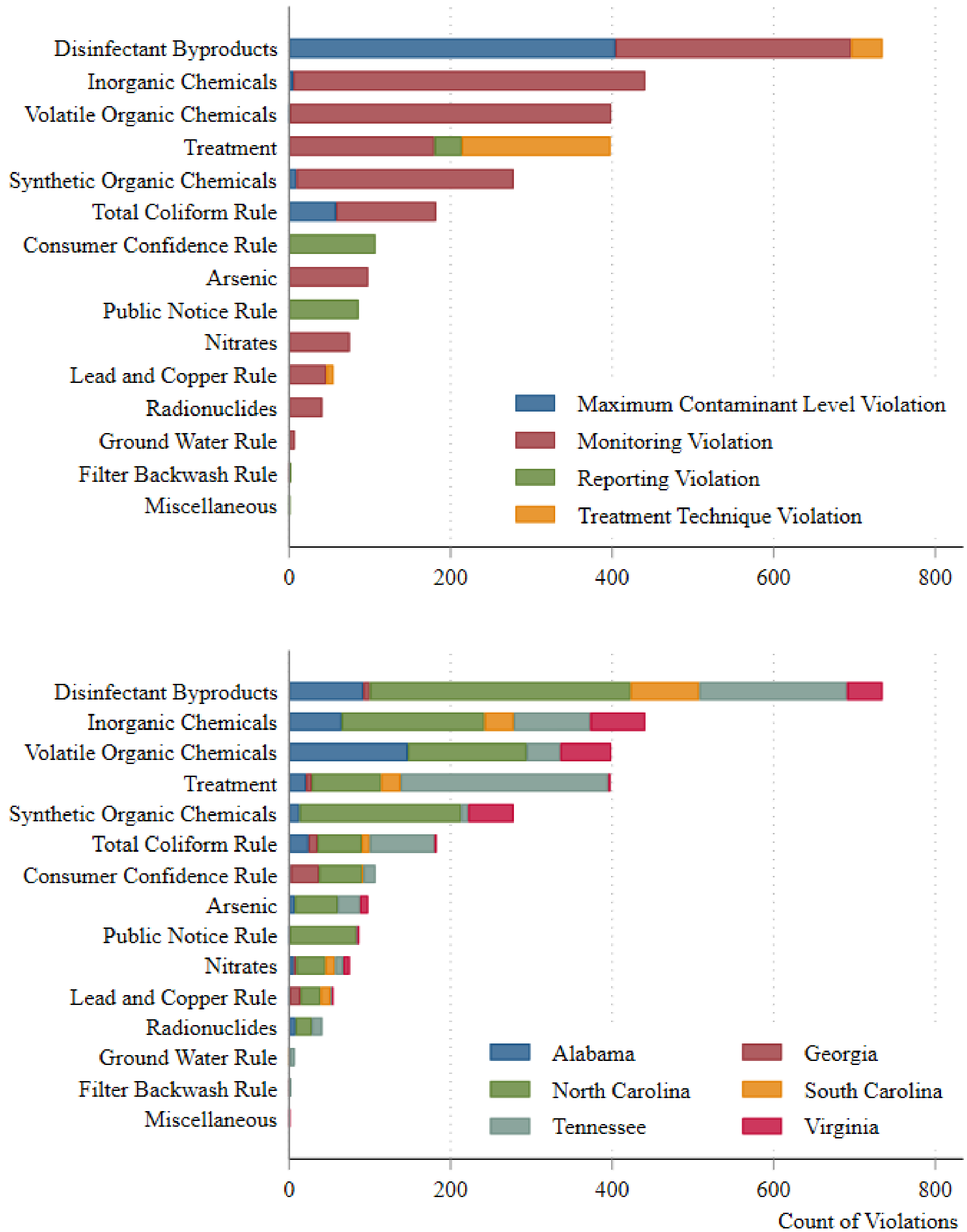
Notes: Average annual violation rate plotted across all water systems excluding transient non-community water systems. The "Downstream" category includes water systems sourcing from coal ash affected waters according to the Southern Environmental Law Center. "Not Downstream" water systems are all other active water systems.

Figure A.9: Safe Drinking Water Act Violations by Rule (2000-2018)



Notes: Average annual violation rate plotted across all water systems excluding transient non-community water systems. The "Downstream" category includes water systems sourcing from coal ash affected waters according to the Southern Environmental Law Center. "Not Downstream" water systems are all other active water systems. Filter backwash, radiation, groundwater, and synthetic organic chemical rules not included. Note that Y axes are not constant across rule names.

Figure A.10: Safe Drinking Water Act Violations in Municipal Water Systems Downstream from Coal Ash Sites (1980-2018)

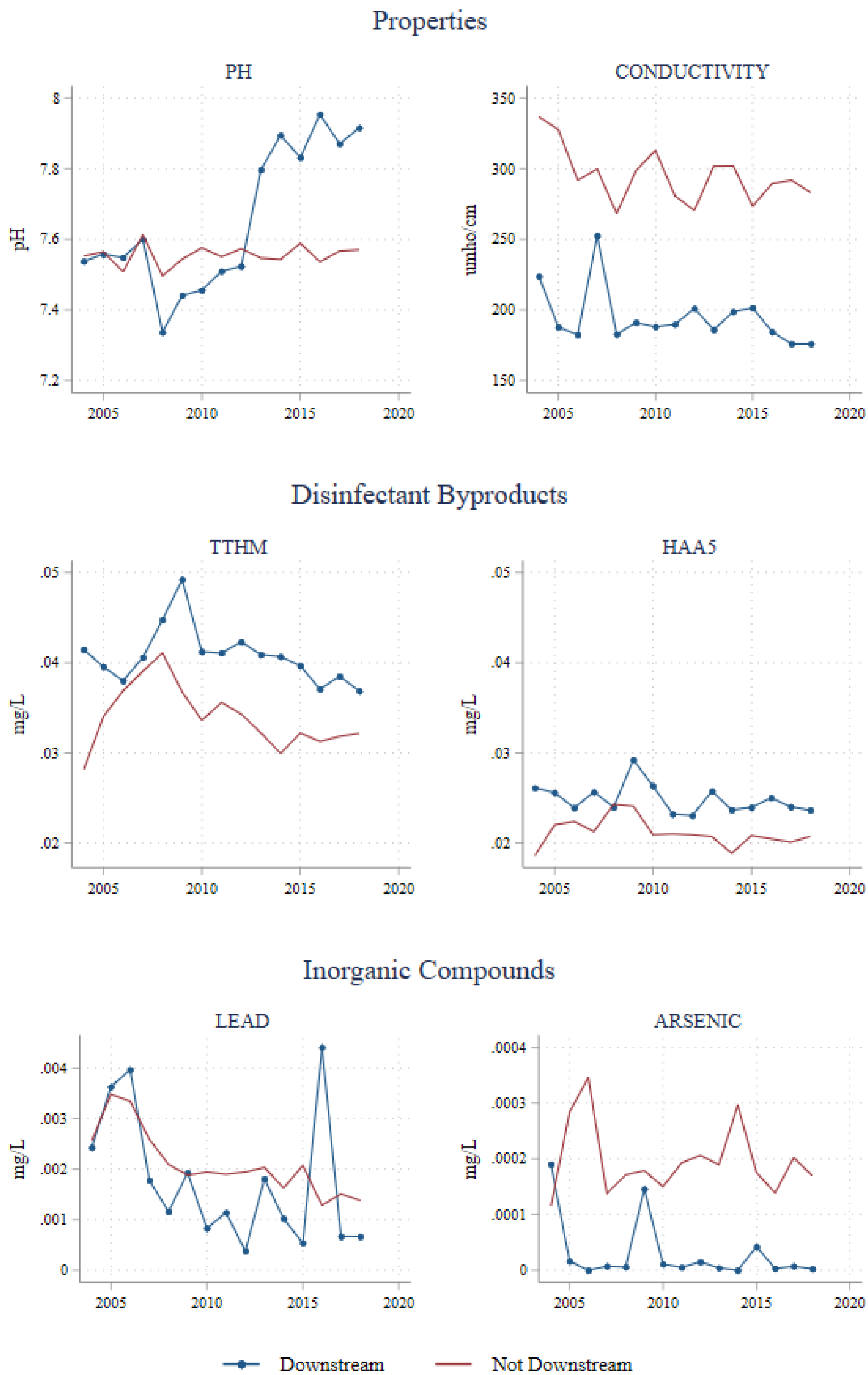


Notes: Each row represents the count of Safe Drinking Water Act violations for any given rule, where the rules are listed down the y-axis. Only municipal water systems designated to be influenced by coal ash according to the Southern Environmental Law Center are included. The top panel provides a breakdown by type of infraction, while the bottom panel shows the state-level burden of these violations.

Figure A.11: Safe Drinking Water Act Violations in Municipal Water Systems Downstream from Coal Ash Sites Over Time (1980-2018)

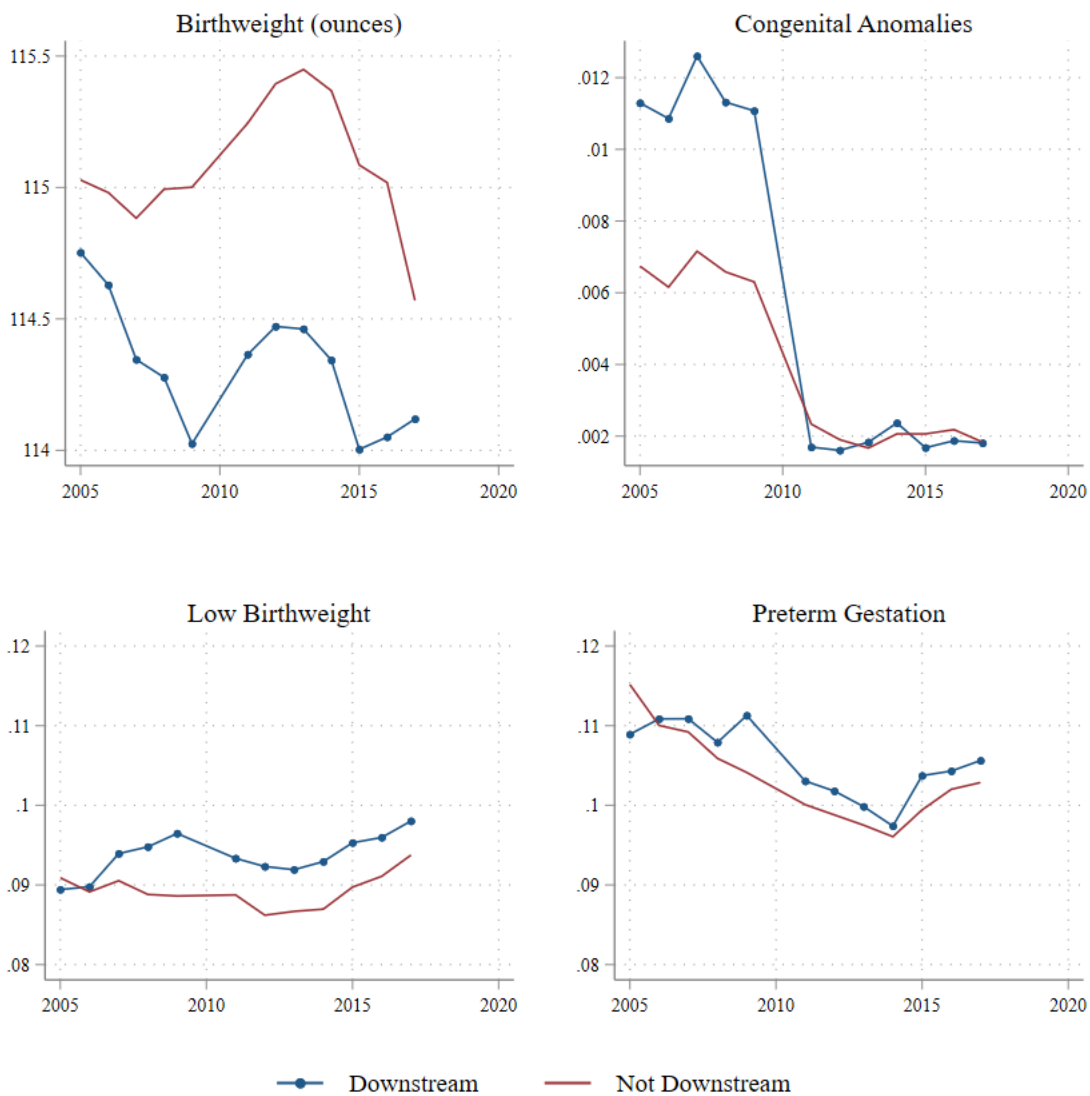
Notes: Each dot represents the count of violations in the given category in a year. Only water systems sourcing from coal ash affected waters according to the Southern Environmental Law Center are included. The top panel provides a breakdown by type of infraction, while the bottom panel shows the state-level burden of these violations.

Figure A.12: Municipal Water Quality Criteria (2005-2018)



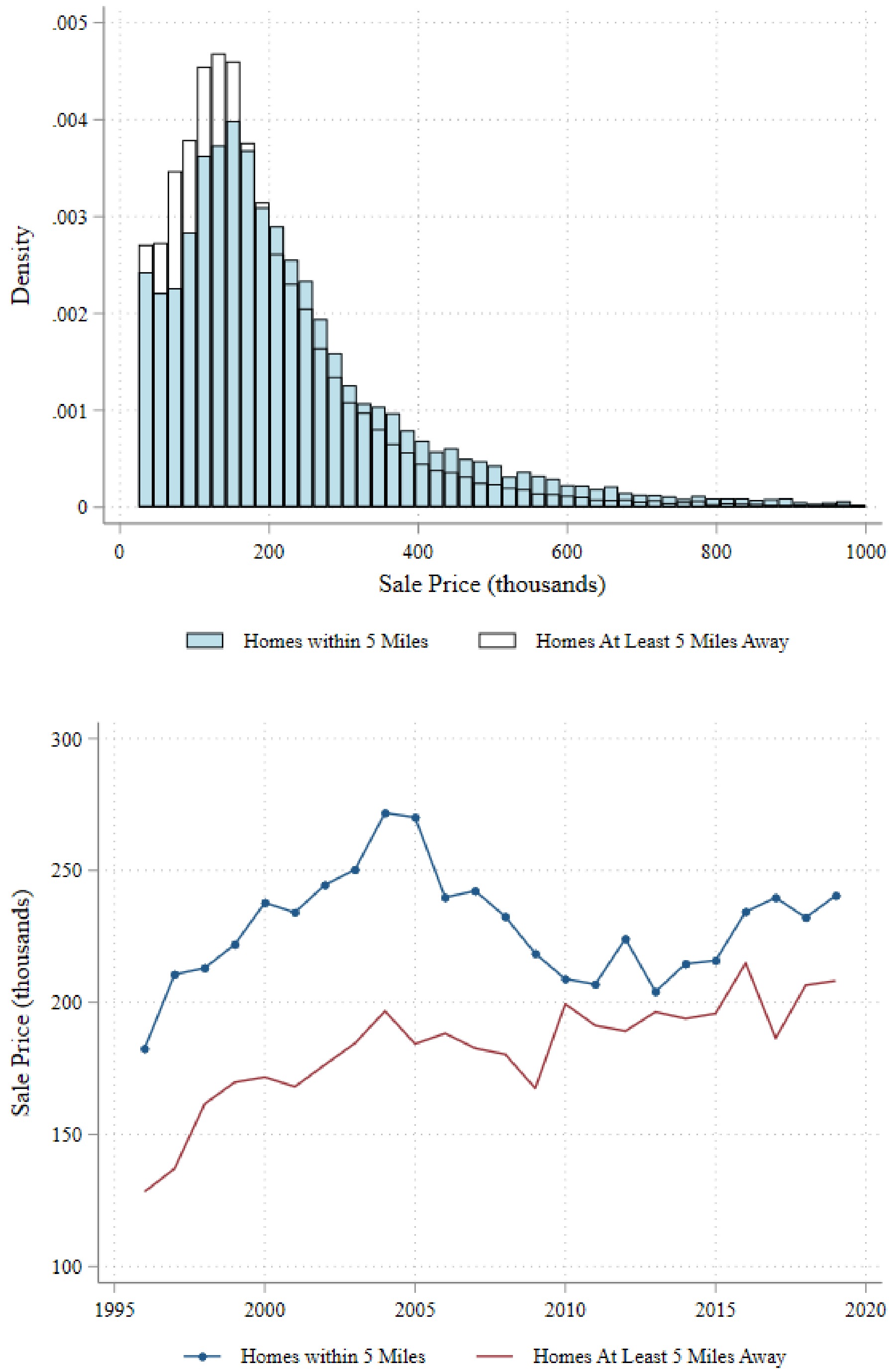
Notes: Average value calculated across all water system sample tests in Alabama, Georgia, North Carolina, South Carolina, and Virginia (i.e., exlcluding Tennessee). The "Downstream" category includes water systems sourcing from coal ash affected waters according to the Southern Environmental Law Center. "Not Downstream" water systems are all other active water systems. Municipal water systems sourcing from surface and groundwater are included..

Figure A.13: Fetal Health Indicators in North Carolina (2005-2017)



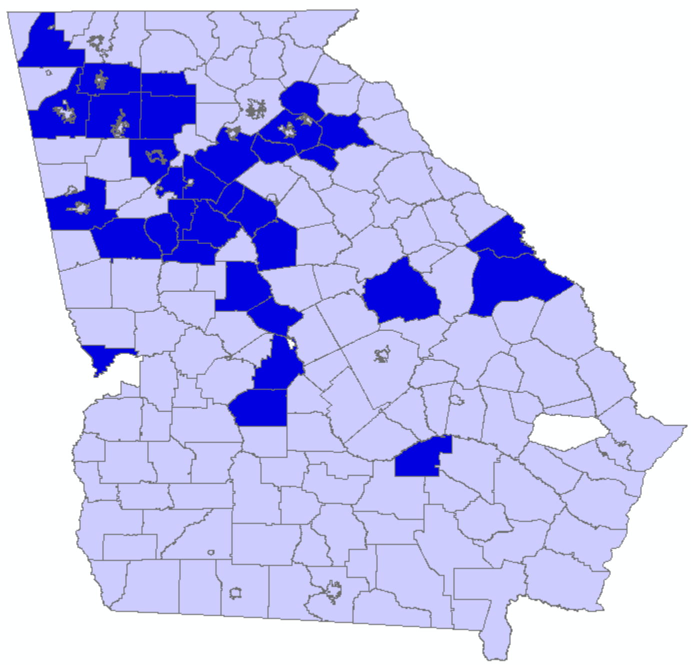
Notes: The "Downstream" category includes mothers ever known to live in service zones of municipal water systems using coal ash affected source waters according to the Southern Environmental Law Center. "Not Downstream" represents fetal health outcomes of all other mothers. Low birthweight is the rate of all newborns born weighing less than 2500 grams. Preterm gestation represents newborns born with estimated gestation length of less than 37 weeks. Congenital anomalies include all fetal abnormalities except chromosomal disorders. The sharp discontinuity in congenital anomalies in 2010 is due to a change in recording practices in that year. In the pre-2010 forms, practitioners recorded a wider variety of conditions on the regular birth form. After the change to the new form, a smaller subset of conditions are reported.

Figure A.14: Home Sale Prices in Counties with Coal Ash Ponds (1996-2019)



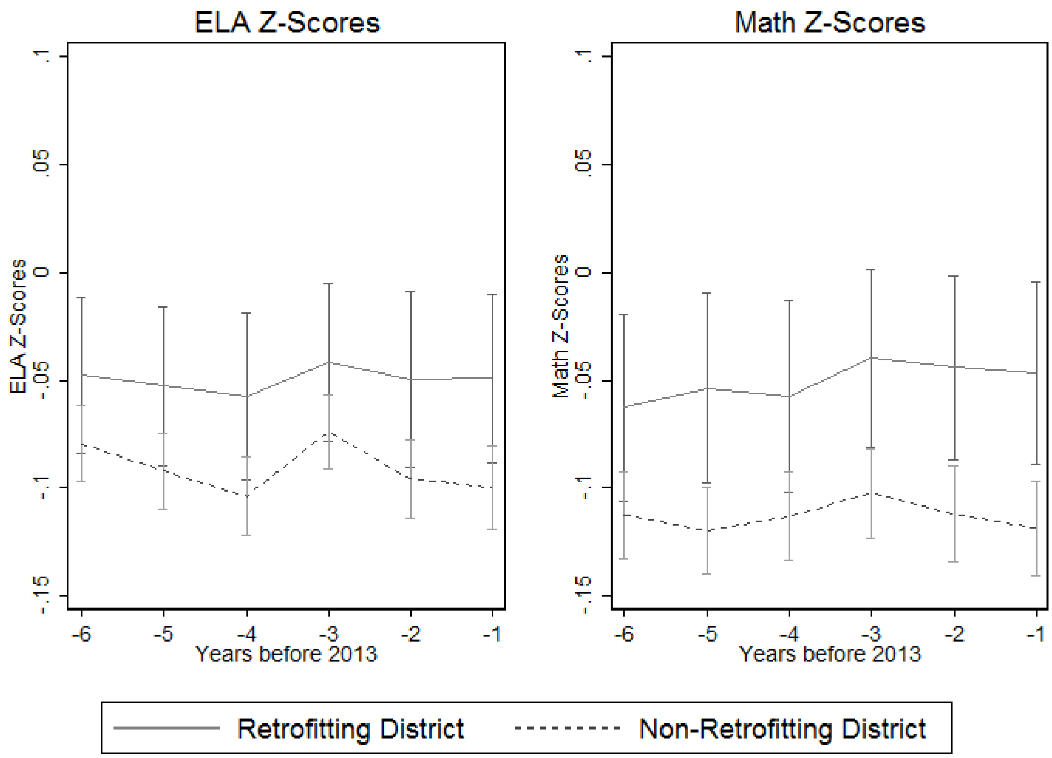
Notes: Homes with sale prices over $1m are excluded from both panels. Certain counties do not have sale information before 2009, leading to the sharp change in that year. Counties with no homes within five miles of a coal ash pond are not included.

Figure A.15: Retrofitting School Districts



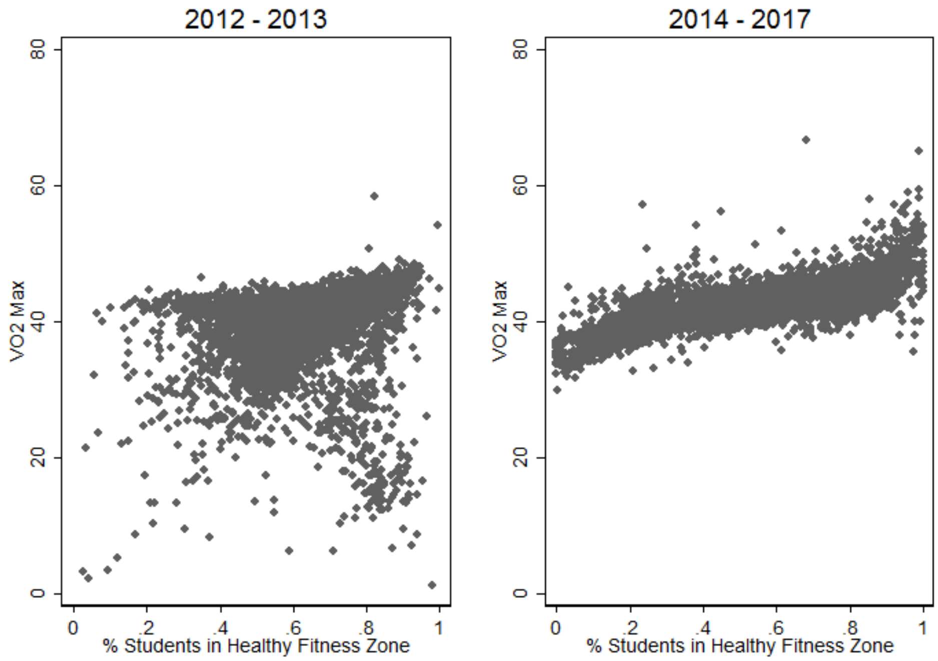
Notes: Darker blue school districts have at least one retrofit cycle during the relevant sample window (2007-2017). Blank districts are missing data.

Figure A.16: Differential Pre-Trends by Retrofitting Districts 2007-2013



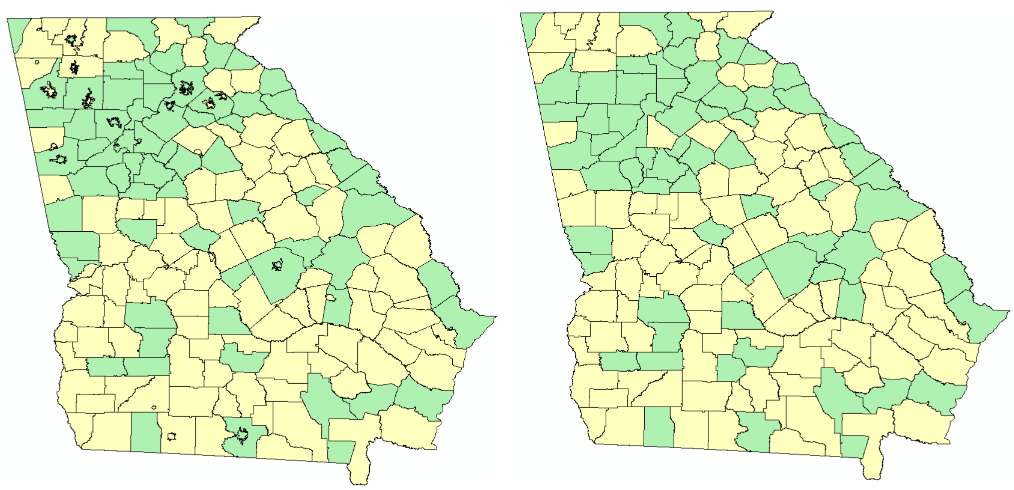
Notes: Figure plots the trend in ELA and Math test scores across retrofitting and non-retrofitting districts before 2013, the mode year of retrofit implementation, such that -1 represents school year 2011-12. Because the timing of retrofits varied across districts, we are unable to conduct a simple event study with non-retrofitting districts as a comparison. We therefore normalize treatment to 2013 and plot trends across districts that ever retrofit and those that do not. Some of the pre-2013 years feature retrofits, and therefore may be expected to have differential trends over this period. Nevertheless, the trend lines are roughly parallel over this sample window.

Figure A.17: Data Issues in Aerobic Capacity



Notes: Each pane scatters the school-level average VO_2 max against the percent of students attaining healthy fitness zone (HFZ) status. The left pane presents the scatterplot for school years 2011-12 and 2012-13, while the right pane displays a scatterplot for the remaining years in the sample. A school's average VO_2 max should be highly correlated with the percent of students attaining HFZ status because each child's VO_2 max is used to determine whether they meet HFZ standards. In the right panel, we observe such a tight relationship between these related measures. In the left panel, however, the relationship is less clear. After the 2012-13 school year, there are no female school-level VO_2 max observations below 26 or male school-level VO_2 max observations below 30. These values are nevertheless very common in the first two years of the sample, and many of these low values correspond to relatively high HFZ attainment. Such values suggest a data-reporting issue in the roll-out years of the sample.

Figure A.18: Farm-to-School Policy Adoption

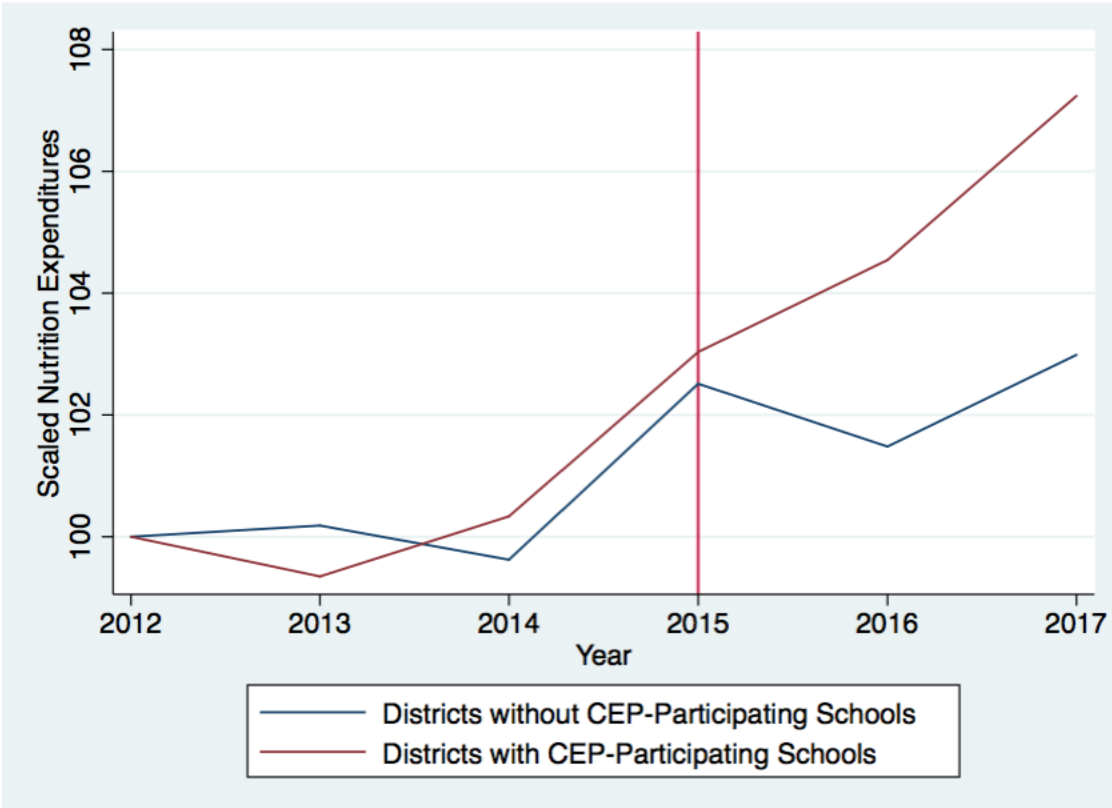


(a) Farm-to-school Districts

(b) Farm-to-school Counties

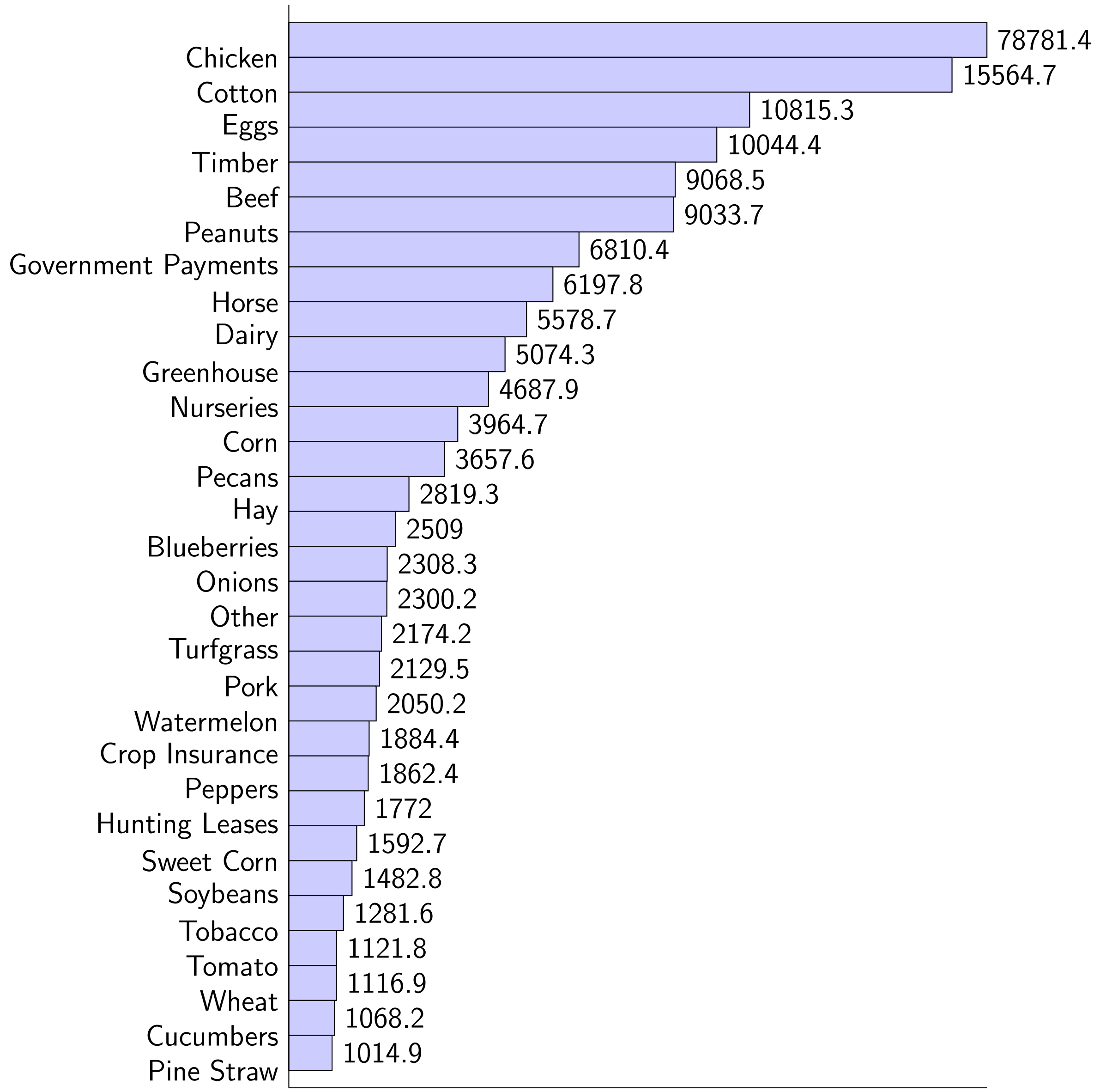
Note: Beige represents school districts that never adopt a farm-to-school policy or counties with no school district adopting a farm-to-school policy. Green regions represent school districts or counties with farm-to-school policies.

Figure A.19: Nutrition Expenditures in Georgia 2012-2017



Note: District nutrition expenditures are scaled such that 2012 expenditures are equal to 100. Lines represent averages over all school districts with CEP-participating schools or without CEP-participating schools. All values expressed in 2017 dollars. The vertical red line represents the last year in which no school participated in the CEP program.

Figure A.20: Total Agricultural Revenues by Product (in millions) 2001-2017



Note: Figure plots all-time revenues by product in the Farm Gate Values Survey in millions of dollars.

Appendix B

Tables

Table B.1: The Quantity of Coal Ash Released by Facility and Type of Compound (2005-2017)

	Mean	SD	
Facility Containment Information			% Missing
Total Ponds	6.73	(3.52)	0%
Average Acres per Pond	85.52	(112.29)	69%
Height *(ft)*	50.58	(42.67)	71%
Lining	0.31	(0.46)	37%
Leachate	0.22	(0.41)	36%
Average Total Coal Ash Production by Plant (tons)			
Heavy Metals	2,717.59	(3066.78)	
Carcinogenic Compounds	248.04	(297.9)	
Quantity Impounded	6,002.58	(7,646.9)	
Surface Water Releases	173.5	(393.6)	
All–Time Surface–Water Releases by Compound (tons)			RSEI Toxixity
Ammonia	11.9	(25.8)	NA
Antimony	1.8	(8.4)	1300
Arsenic	166.4	(249.9)	3000
Barium	3524.9	(5486.1)	2.5
Beryllium	14.8	(30.6)	250
Chromium	290.2	(317.9)	170
Cobalt	99.6	(143.9)	NA
Copper	359.3	(388.0)	750
Lead	155.8	(161.1)	8800
Manganese	509.6	(534.7)	3.6
Mercury	0.002	(0.003)	5000
Nickel	253.0	(280.1)	10
Nitrate	45.7	(264.9)	0.31
Selenium	16.5	(35.5)	100
Thallium	19.3	(66.8)	7100
Vanadium	632.7	(652.6)	71
Zinc	404.2	(449.8)	1.7
Plant-Year Observations with Positive Releases	526		
Steam–Generating Coal Electricity Plants	63		

Mean coefficients reported; standard deviations in parentheses. Observations in the second panel are at the plant level, reflecting totals across all plants in all years from 2005-2017. The third panel displays average sum of all surface water releases by compound across pollution release sites.

Table B.2: Analyte Testing, Violation Rates, and Water System Characteristics 2005-2017

	Within 25 Miles Downstream		Not Within 25 Miles Downstream	
Surface Water Monitors (2005-2017)				
Arsenic (mg/l)	0.3958	(1.8176)	0.7785	(6.877)
Chromium (mg/l)	1.9103	(8.9721)	2.7691	(15.1431)
Conductivity (us/cm)	8994.3	(14089.9)	5030.7	(11422.4)
Dissolved Oxygen (mg/l)	5.073	(2.688)	7.393	(24.506)
Lead (mg/l)	1.0357	(4.8987)	3.6671	(50.27)
PH	7.32	(0.605)	7.27	(0.753)
Selenium (mg/l)	0.1218	(0.7242)	0.1115	(0.5329)
Temperature (c)	24.310	(7.598)	19.639	(12.640)
Monitor Observations	748,988		4,848,838	
Monitors	2,064		122,163	
Municipal Water Systems				
Service Population (thousands)	50.732	(97.585)	2.308	(19.691)
Service Connections (thousands)	20.412	(40.155)	0.517	(5.46)
Age in 2018	35.62	(6.34)	27.334	(11.72)
State Regulatory Monitoring Tests (2005-2017)				
Arsenic (mg/l)	0.00002	(0.0005)	0.0020	(0.4723)
Conductivity (us/cm)	183.44	(264.4)	299.10	(1012.6)
Lead (mg/l)	0.0017	(0.0309)	0.0058	(2.423)
Haloacetic Acids (mg/l)	0.0246	(0.0150)	0.0228	(0.4034)
PH	7.796	(.6041)	7.725	(0.6806)
Trihalomethanes (mg/l)	0.0417	(0.0219)	0.0359	(0.4431)
Safe Drinking Water Inventory System Violations (2000-2018)				
Total Violations	10.396	(14.781)	7.996	(28.322)
Health-Based Violations	2.734	(4.145)	0.7357	(2.9364)
Annual Violation Rate	0.1670	(0.3730)	0.1285	(0.3347)
Health-based Violation Rate	0.0698	(0.2549)	0.0225	(0.1482)
Maximum Contaminant Level	0.0511	(0.2201)	0.0197	(0.1390)
Monitoring Violation	0.0901	(0.2864)	0.0935	(0.2912)
Reporting Violation Rate	0.0344	(0.1822)	0.0371	(0.1890)
Treatement Technique	0.0219	(0.1463)	0.0029	(0.0542)
Arsenic	0.0047	(0.0683)	0.0014	(0.0374)
Consumer Confidence Rule	0.0279	(0.2092)	0.0218	(0.2185)
Disinfectant Byproducts	0.1771	(0.7811)	0.0308	(0.3496)
Inorganic Compounds	0.0477	(0.7468)	0.0165	(0.4333)
Lead and Copper	0.0109	(0.1287)	0.0163	(0.1911)
Public Notice	0.0224	(0.2824)	0.0603	(0.6036)
Volatile Organic Chemicals	0.0711	(1.3958)	0.0688	(1.6584)
Water System Samples	162,790		1,185,225	
Water System Years	42,722		491,892	
Water Systems	193		3,839	

Mean coefficients reported; standard deviations in parentheses. Observations are at the water system and water-system-year level. Surface monitor sample restricted to samples in streams, lakes, or rivers. Observations include only monitors reporting results for arsenic, chromium, conductivity, dissolved oxygen, lead, pH, selenium, and temperature. Municipal water system sample excludes transient non-community water systems. Sample time window is 2005-2017 for surface water and municipal monitoring information and 2000-2018 for Safe-Drinking Water Inventory System (SDWIS) violation reports.

Table B.3: Mother, Birth, and Home Sale Information in Potentially Affected and Unaffected Regions

	Ever Served by Affected Municipal Water System		Never Served by Affected Municipal Water System	
Mother Characteristics (2005-2017)				
Age	27.58	(5.99)	27.54	(6.01)
Asian	0.042	(0.201)	0.031	(0.173)
Black	0.303	(0.459)	0.212	(0.409)
Hispanic	0.161	(0.367)	0.155	(0.362)
White	0.552	(0.497)	0.656	(0.478)
Married	0.567	(0.495)	0.604	(0.489)
HS diploma or Less	0.424	(0.494)	0.443	(0.496)
Prenatal Visits	11.86	(4.27)	12.20	(4.23)
Tobacco	0.089	(0.286)	0.104	(0.305)
Hypertension*	0.049	(0.217)	0.044	(0.204)
Diabetes*	0.039	(0.194)	0.036	(0.187)
Birth Characteristics (2005-2017)				
Ounces	114.32	(21.82)	115.08	(21.84)
Low Birthweight (2500 grams)	0.094	(0.291)	0.089	(0.285)
Preterm Gestation (37 weeks)	0.106	(0.307)	0.103	(0.304)
Congenital Anomalies	0.005	(0.069)	0.003	(0.053)
Female	0.489	(0.499)	0.488	(0.499)
Movers	0.150	(0.357)	0.098	(0.298)
PM 2.5 Mean	10.49	(2.29)	9.97	(2.28)
PM 2.5 Max	16.32	(4.84)	15.97	(5.03)
Birth Observations	356,868		1,101,204	
Unique Mothers	241,188		779,974	
	Homes Within 5 Miles of Ash Pond		Homes Not Within 5 Miles of Ash Pond	
Properties and Sales (1996-2018)				
Average Sale Value (thousands)	228.1	(201.2)	192.7	(163.1)
Avg. No. Sales	1.537	(0.938)	1.590	(0.985)
Lotsize (thousands sq ft.)	50.6	(351.5)	110.6	(1,080.0)
Bedrooms	2.797	(1.289)	2.678	(1.615)
Baths	1.811	(0.999)	1.753	(1.231)
Home Sales	37,224		248,743	
Unique Homes	24,699		157,000	

Mean coefficients reported; standard deviations in parentheses. Sample of mothers includes all residents in the state, including those not assigned to municipal water service zones. Sample of home sales limited to 12 counties with a coal ash containment facility. *Refers to the gestational diabetes or pre-existing diabetes and gestational hypertension or pre-existing hypertension.

Table B.4: Water Quality Indicators of Surface Waters Downstream from Coal Ash Sites (2005-2017)

	Downstream (1)	Releases (binary) (2)	Releases (continuous) (3)
Inorganic Compounds			
Arsenic	0.0863**	0.0576	0.0021
Dep. Var. Mean = 0.4596	(0.0373)	(0.0366)	(0.0022)
Observations	[36,715]	[36,715]	[36,715]
Chromium	0.1538	-0.0313	-0.0018*
Dep. Var. Mean = 1.627	(0.3353)	(0.0757)	(0.0007)
Observations	[57,089]	[57,089]	[57,089]
Lead	0.1730	0.4992	-0.0124***
Dep. Var. Mean = 1.516	(0.1662)	(0.3538)	(0.0020)
Observations	[61,731]	[61,731]	[61,731]
Selenium	0.0190***	0.0179***	0.0008*
Dep. Var. Mean = 0.0536	(0.0066)	(0.0020)	(0.0005)
Observations	[28,928]	[28,928]	[28,928]
Properties			
Conductivity	1567.42	-333.19**	1.050
Dep. Var. Mean = 5279.45	(1932.85)	(147.58)	(3.077)
Observations	[1,119,939]	[1,119,939]	[1,119,939]
Dissolved Oxygen	-0.6367**	0.0237	-0.0006
Dep. Var. Mean = 6.982	(0.2491)	(0.0362)	(0.0011)
Observations	[1,097,515]	[1,097,515]	[1,097,515]
pH	0.1948***	0.0464**	0.0007
Dep. Var. Mean = 7.28	(0.1384)	(0.0174)	(0.0011)
Observations	[1,227,668]	[1,227,668]	[1,227,668]
Temperature	1.0293***	-0.0435	-0.0009*
Dep. Var. Mean = 20.275	(0.0407)	(0.0407)	(0.0006)
Observations	[1,240,357]	[1,240,357]	[1,240,357]
Monitor		✓	✓
Watershed-by-Year	✓	✓	✓
Watershed-by-Year	✓	✓	✓

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors two-way clustered at the monitor and watershed in parentheses. The first column regresses an indicator for whether a monitor is within 25 miles downstream of a coal ash release site on the level of an analyte depicted in the row title. Column (2) regresses an indicator for whether coal ash is released upstream of a monitor in year t on the compound's concentration. Column (3) regresses a continuous measure of the sum of coal ash released upstream in any given year on the water quality indicator. Controls include a dummy for abnormal weather events, dummy indicators for the hydrologic conditions of the river system, and dummy indicators for the sample medium (e.g., sediment or surface water). Analytic sample weights included. All regressions performed assuming coal ash influence cutoff distance of 25 miles (40 kilometers). Note mean analyte levels may differ from figures because analytes of different media are included in the regressions with corresponding controls.

Table B.5: Water Quality Indicators of Municipal Waters Downstream from Coal Ash Sites (2005-2017)

	Downstream (1)	Releases Binary (2)	Annual Tons Released (3)
Disinfectant Byproducts			
Haloacetic Acids (HAA5)	-0.0026*	-0.0032	-0.0001
Dep. Var. Mean= 0.0220	(0.0010)	(0.0047)	(0.0001)
Observations	[249,467]	[249,467]	[249,467]
Trihalomethanes (TTHM)	0.0007	-0.0099	-0.0003
Dep. Var. Mean= 0.0362	(0.0030)	(0.0088)	(0.0002)
Observations	[249,132]	[249,132]	[249,132]
Inorganic Compounds			
Arsenic	-0.0058	0.0084	0.0007
Dep. Var. Mean= 0.0027	(0.0075)	(0.0123)	(0.0009)
Observations	[46,729]	[46,729]	[46,729]
Lead	0.0081	-0.0033	0.0035***
Dep. Var. Mean= 0.0070	(0.0089)	(0.0014)	(0.0003)
Observations	[364,643]	[364,643]	[364,643]
Properties			
Conductivity	-120.43	45.99**	3.37***
Dep. Var. Mean = 291.00	(75.03)	(19.10)	(1.08)
Observations	[29,697]	[29,697]	([29,697]
pH	-0.3765**	-0.0172***	0.0070**
Dep. Var. Mean= 7.76	(0.0427)	(0.0001)	(0.0008)
Observations	[71,059]	[71,059]	[71,059]
Water System		✓	✓
State-by-Year	✓	✓	✓
Month	✓	✓	✓
Watershed	✓		

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors two-way clustered at the municipal water system and state in parentheses. The first column regresses an indicator for whether a municipal water system sources from coal ash affected waters according to the Southern Environmental Law Center. Column (2) regresses an indicator for whether coal ash is released upstream of a municipal water system's intake in year t on the compound's concentration. Column (3) regresses a continuous measure of the sum of coal ash released upstream in any given year on the water quality indicator. Transient non-community water systems are excluded, as are any water systems with fewer than three tests of the given water quality analyte over the sample period. Controls include dummies for the facility type where the test occurred, system size dummies, system age, and a dummy variable equal to one if the analyte was not detected. Analytic sampling weights included.

Table B.6: Upstream Coal Pollution and the Probability of a Water Quality Violation (2000-2018)

	Time-Varying Binary Coal Ash Releases		Time-Varying Continuous Coal Ash Releases	
	(1)	(2)	(3)	(4)
	β	dy/dx	β	dy/dx
Violations by Infraction Type				
Any Violation	-0.0370	-0.0045	-0.0002	-0.0005
Dep. Var. Mean = 0.1278	(0.0279)	(0.0183)	(0.0002)	(0.0004)
Health-based Violation	-0.0035	0.0121**	-0.0001	-0.0000
Dep. Var. Mean = 0.0228	(0.0148)	(0.0062)	(0.0001)	(0.0001)
Maximum Contaminant Level	-0.0004	0.0103*	-0.0000	-0.0000
Dep. Var. Mean = 0.0199	(0.0138)	(0.006)	(0.0000)	(0.0001)
Monitoring Violation	-0.0581*	-0.0165	-0.0001	-0.0001
Dep. Var. Mean = 0.0935	(0.0252)	(0.0164)	(0.0002)	(0.0002)
Reporting Violation	0.0320*	0.0250**	-0.0000	-0.0001
Dep. Var. Mean = 0.0370	(0.0173)	(0.0106)	(.0001)	(.0001)
Violations by Rule Type				
Arsenic	0.0078**	0.0034***	2.77e-06	8.80e-06
Dep. Var. Mean = 0.0014	(0.0040)	(0.000)	(0.0002)	(6.91e-06)
Disinfectant Byproducts	-0.0071	0.0033	0.0010**	-0.0001
Dep. Var. Mean = 0.0134	(0.012)	(0.0038)	(0.0001)	(0.0000)
Inorganic Compounds	0.0088**	0.0039***	0.0015**	-4.66e-06
Dep. Var. Mean = 0.0024	(0.0045)	(0.000)	(0.0000)	(9.52e-06)
Lead and Copper	-0.0077	-0.0109**	-9.85e-06	-0.0003
Dep. Var. Mean = .0112	(0.0103)	(0.0000)	(0.0004)	(0.0003)
Volatile Organic Compounds	0.0015	0.0012	2.33e-06	8.71e-06
Dep. Var. Mean = 0.0024	(0.0048)	(0.0021)	(0.0000)	(0.0000)
Observations	247,794	247,794	247,794	247,794
Water Systems	15,493	15,493	15,493	15,493

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the water system in parentheses. Standard error of the marginal effect dy/dx calculated using the delta method. Dependent variable means are the average of all active water system-year combinations, where a water system-year is equal to one if the water system experienced a violation of the specified type and zero otherwise. Time-varying binary coal ash releases is equal to one if a municipal water system was potentially affected by any coal ash releases in a given year and zero otherwise. Time-varying continuous coal ash releases is equal to the tons of coal ash released within 25 miles upstream and zero otherwise. In the probit model, controls include system size dummies, federal water system type (e.g., community water system), owner type, school water system, surface water-sourced system, protected source water system, and age of the municipal water system.

Table B.7: CCRs, Water Quality, and Fetal Health 2005-2017

	Full Sample				Mothers with a High School Degree or Less			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Birthweight (ozs)	Low Birthweight	Preterm Gestation	Congenital Anomalies	Birthweight (ozs)	Low Birthweight	Preterm Gestation	Congenital Anomalies
Panel A.								
Downstream	-1.2411***	0.0171***	0.0126*	0.0032	-2.2118***	0.0286***	0.0255**	0.0021
	(0.4127)	(0.0065)	(0.0020)	(0.0017)	(0.6251)	(0.0107)	(0.0109)	(0.0034)
PM2.5	-0.949***	0.0111***	0.0195***	0.0001	-1.001**	0.0126***	0.0230***	0.0002
	(0.0514)	(0.0007)	(0.0012)	(0.0002)	(0.0794)	(0.0012)	(0.0012)	(0.0003)
Panel B.								
Releases (binary)	0.4147***	-0.0029	-0.0043**	-0.0007	0.8128***	-0.0068**	-0.0087**	-0.0014*
	(0.1241)	(0.0021)	(0.0020)	(0.0004)	(0.2023)	(0.0033)	(0.0034)	(0.0006)
PM2.5	-0.954***	0.0111***	0.0195***	0.0001	-1.018***	0.0126***	0.023***	0.0002
	(0.0514)	(0.0007)	(0.0008)	(0.0002)	(0.0794)	(0.0012)	(0.0013)	(0.0003)
Panel C.								
Releases (continuous)	0.0143	-0.0002	-0.0088**	-0.0001	0.0161	-0.0004	-0.0002	-0.0001*
	(0.0127)	(0.0002)	(0.003)	(0.0002)	(0.0184)	(0.0005)	(0.0003)	(0.0001)
PM2.5	-0.952	0.0111***	0.0231***	0.0001	-1.012***	0.0195***	0.0231***	0.0002
	(0.0514)	(0.0007)	(0.0013)	(0.0008)	(0.0794)	(0.0012)	(0.0018)	0.0003
Mother Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Zipcode Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Dep. Var. Mean	114.89	0.0903	0.1040	0.0044	114.89	0.0903	0.1040	0.0044
Observations	747,468	747,468	747,468	747,468	303,110	303,110	303,110	303,110

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the mother in parentheses. Mother and zipcode fixed effects included. Low birthweight refers to births of less than 2500 grams. Preterm gestation represents a birth with gestation of less than 37 weeks. The congenital anomaly indicator excludes chromosomal anomalies and trisomy 21. Mean PM 2.5 represents the average PM 2.5 concentration in the mother's county of residence over the entire gestational period. Additional controls include maximum and maximum PM 2.5 squared in the county of residence across all months of gestation, gender of the newborn, dummies for birth order, mother's age at time of birth, mother's age squared, six dummy bins for number of clinic visits during gestation, an indicator for tobacco use during gestation, and an indicator for having moved since the last preganancy.

Table B.8: CCRs and Fetal Health by In- and Out-Movers 2005-2017

	(1) Birthweight (ozs)	(2) Low Birthweight	(3) Preterm Gestation	(4) Congenital Anomalies
In Movers (=1)	-1.8378***	0.0280***	0.0211**	-0.0001
	(0.4419)	(0.0069)	(0.0073)	(0.0022)
Out Movers (=1)	0.5801	-0.0100	-0.0023	-0.0055**
	(0.4342)	(0.0068)	(0.0072)	(0.0021)
PM2.5	-0.9502***	0.0111***	0.0196***	0.0001
	(0.051)	(0.0007)	(0.0008)	(0.0002)
Mother Fixed Effects	✓	✓	✓	✓
Zipcode Fixed Effects	✓	✓	✓	✓
Dep. Var. Mean	114.89	0.0903	0.1040	0.0044
Observations	747,468	747,468	747,468	747,468

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the mother in parentheses. Mother fixed effects included. Low birthweight refers to births of less than 2500 grams. Preterm gestation represents a birth with gestation of less than 37 weeks. Mean PM 2.5 represents the average PM2.5 concentration in the mother's county of residence over the entire gestational period. Additional controls include maximum and maximum PM 2.5 squared in the county of residence across all months of gestation, gender of the newborn, dummies for birth order, mother's age at time of birth, mother's age squared, six dummy bins for number of clinic visits during gestation, an indicator for tobacco use during gestation, and an indicator for having moved since the last preganancy.

Table B.9: How Mandatory House Well Testing Affected House Sale Values After the Coal Ash Management Act of 2014

Distance Cutoff	(1) 1 Mile	(2) 2.5 Miles	(3) 5 Miles	(4) 1 Miles	(5) 2.5 Miles	(6) 5 Miles
Near*Post	-45,295.4***	-36,406.9***	-24,691.8***	-37,333.5***	-16,090.1***	-12,673.9***
	(17,403.2)	(5,151.2)	(2,371.5)	(12,591.3)	(2,784.1)	(2,229.5)
Mean Sale Price	320,307.6	259,978.8	248,597.3	320,307.6	259,978.8	248,597.3
% Change	-14.1	-13.9	-9.7	-11.6	-6.1	-4.8
Δ Total House Value	-24.4M	-180.2M	-448.2M	-19.9M	-79.6M	-228.7M
City and Year FEs	✓	✓	✓			
House and Year FEs				✓	✓	✓
Home Sales in Sample	226,973	226,973	226,973	163,077	163,077	163,077
Unique Homes	181,669	181,669	181,669	63,963	63,963	63,963
Affected Home Sales	294	2,990	13,540	308	2,238	8,377

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the county in parentheses. The dependent variable is house sale price. The independent variable is the interaction of being within the specified distance of a coal ash pond and an indicator for sales occurring after 2014. Total change in home value is the product of the change in home values and the number of sales after 2014, where the number of sales is 538, 4950, and 18,154, ordered by distance cutoff. Regressions (1) to (3) may have more or fewer observations than (4) to (6) because many homes are not incorporated into cities. The counts of affected homes, unique homes, and sales reflect the number of sales in the regression sample rather than the total number of sales. Sample excludes home sales with valuation in excess of $1.5 million. The Coal Ash Management Act mandated testing drinking wells of homes within 2,500 feet of ash ponds, leading to information disclosure that over 97% of homes had been using well-water considered unsafe to drink by the EPA.

Table B.10: Surface Water Monitoring Tests in the Water Quality Portal (2005-2017)

Constituent *(units)*	N	%BDL	Min	Median	Max	Monitors	Watersheds
Aluminum (mg/kg)	110,768	20.21	0	0.102	120,000	5402	230
Antimony (mg/kg)	40,714	58.09	0	0.001	55	3998	199
Arsenic (mg/kg)	107,107	53.61	0	0.001	430	5959	232
Beryllium (mg/kg)	50,839	69.54	0	0.0003	55	2785	160
Bromide (mg/kg)	10,064	20.21	0	0.038	60.3	448	70
Cadmium (mg/kg)	151,379	71.72	0	0.0005	1100	7821	236
Calcium (mg/kg)	104,525	4.35	0	7.8	52000	6026	234
Chemical oxygen demand (mg/kg)	15,366	15.78	0	7.8	1700	740	102
Chromium (mg/kg)	147,469	65.43	0	0.001	970	7615	236
Conductivity (uS/cm)	2,237,496	0.22	-2.47	167	511170	20629	239
Copper (mg/kg)	175,735	61.93	0	0.002	3100	8065	236
Fixed suspended solids (mg/kg)	104,996	4.40	0	8	26067	2435	62
Iron (mg/kg)	192,100	13.50	0	0.339	314000	8185	236
Lead (mg/kg)	156,963	61.56	0	0.001	11000	8015	236
Magnesium (mg/kg)	106,114	4.86	0	2.42	21300	6101	236
Manganese (mg/kg)	191,461	17.03	0	0.048	26000	7904	236
Mercury (mg/kg)	123,183	61.66	0	0.0002	274	7044	234
Nickel (mg/kg)	139,411	61.14	0	0.0258	490	7336	236
Nitrogen (mg/kg)	220,222	11.21	0	0.56	4587	6698	111
pH	2,762,327	0.09	0	7.24	16	21559	240
Phosphorus (mg/kg)	706,766	10.79	0	0.05	8700	17276	238
Selenium (mg/kg)	93,791	64.52	0	0.0007	25	5423	231
Silicon (mg/kg)	93,791	5.38	0	2.490	53.71	223	36
Thallium (mg/kg)	39,476	69.14	0	0.0001	100	3483	177
Titanium (mg/kg)	39,476	74.10	0	0.007	14000	1018	83
Total Coliform (MPN/100 ml)	30,102	7.05	0	2200	2.00e+07	302	46
Total dissolved solids (mg/kg)	173,964	2.24	0	81	1010000	3791	173
Total solids (mg/kg)	78,048	1.16	0	104	151000	2953	133
Total suspended solids (mg/kg)	504,347	15.51	0	9.21	38400	14002	239
Total volatile solids (mg/kg)	56,643	1.77	0	8	18500	2106	67
Trihalomethanes (mg/kg)	5,514	79.09	0.0001	0.0003	4.5	202	32
Turbidity (ntu)	674,007	2.62	-1.6	6.7	7417434	13140	239
Vanadium (mg/kg)	20,468	32.83	0	0.0014	570	1348	129
Volatile suspended solids (mg/kg)	39,862	10.89	0	3.6	1150	408	43
Zinc (mg/kg)	182,069	46.63	0	0.01	4500	8063	236

%BDL is the percent of samples that are below the detection limit.

Table B.11: Additional Chemical Compounds in Surface Waters Downstream from Coal Ash Sites (2005-2017)

	Ever Affected	Releases (binary)		Releases (continuous)	
	(1)	(2)	(3)	(4)	(5)
Antimony	0.0096	0.00149	0.00003	0.00014***	0.00071***
	(0.00169)	(0.0026)	(0.00002)	(0.00004)	(0.00010)
Cadmium	-0.01937	-0.02575	-0.05171*	0.00009	-0.00055
	(0.04562)	(0.05915)	(0.02911)	(0.00039)	(0.00067)
Copper	-0.00021	-0.00012	-0.00055*	0.00052***	0.00165*
	(0.00397)	(0.00481)	(0.00029)	(0.00011)	(0.00091)
Mercury	-0.01709	-0.02371	0.00257*	-0.00071	-0.00092
	(0.01952)	(0.02629)	(0.00139)	(0.00058)	(0.00121)
Thallium	-0.00008	-0.00012	6.80e-06**	5.74e-06	0.000039***
	(0.00019)	(0.00028)	(2.47e-06)	(5.30e-06)	(8.38e-06)
Turbidity	3.3230	.59735	-17.409	0.04202	-0.1099
	(2.6664)	(1.7026)	(18.477)	(0.03597)	(0.2396)
Zinc	0.01830	0.02136	-0.00253	0.00065***	-0.00103
	(0.02008)	(0.02286)	(0.00268)	(0.00014)	(0.00596)
Monitor			✓		✓
Watershed-Year	✓	✓	✓	✓	✓
Watershed-Month	✓	✓	✓	✓	✓

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors two-way clustered at the monitor and watershed in parentheses. The first column regresses an indicator for whether a monitor is within 25 miles downstream of a coal ash release site on the level of an analyte depicted in the row title. Columns (2) and (3) regress an indicator for whether coal ash is released upstream of a monitor in year t on the compound's concentration. Columns (4) and (5) regress a continuous measure of the sum of coal ash released upstream in any given year on the water quality indicator. Controls include a dummy for abnormal weather events, dummy indicators for the hydrologic conditions of the river system, and dummy indicators for the sample medium (e.g., sediment or surface water). Analytic sample weights included. All regressions performed assuming coal ash influence cutoff distance of 25 miles (40 kilometers).

Table B.12: Do Counties with Coal Ash Releases Have More Surface Water Pollution from Other Sources? (2005-2017)

	(1) Tons of Surface Water Pollution	(2) Tons of Impounded Pollution
Coal Plant County (=1)	18.45	177.19
	(33.57)	(140.82)
State Fixed Effects	✓	✓
Year Fixed Effects	✓	✓
Dep. Var. Mean	74.18	101.23
Observations	6,406	6,406

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the state in parentheses. The first column regresses an indicator for whether a county has a coal ash pollution site on the quantity of non-coal ash surface water pollution. The second column regresses an indicator for whether a county has a coal ash site on the quantity of non-coal ash pollution impounded in any landfill.

Table B.13: District-Level Student Characteristics

	(1) Non-Retrofitting Districts		(2) Retrofitting Districts		(3) Difference T-Test of Means	
Health Outcomes (2012-2017)						
Aerobic Capacity ($V0_2$ Max)	41.160	(1.688)	41.201	(1.422)	-0.0412	(-0.12)
Body-Mass Index	21.069	(0.880)	20.633	(0.340)	0.436*	(2.54)
AC Attempts / Enrollment	0.407	(0.114)	0.425	(0.079)	-0.0174	(-0.76)
BMI Attempts / Enrollment	0.654	(0.153)	0.689	(0.108)	-0.0346	(-1.12)
Schooling Outcomes (2007-2017)						
Math Z-Scores	-0.107	(0.263)	-0.060	(0.216)	-0.0473	(-0.88)
ELA Z-Scores	-0.107	(0.229)	-0.061	(0.194)	-0.0459	(-0.98)
Attendance rate	95.573	(0.630)	95.584	(0.488)	-0.0112	(-0.09)
Demographics (2007-2017)						
African American	0.367	(0.272)	0.363	(0.266)	0.004	(0.07)
Hispanic	0.082	(0.105)	0.109	(0.077)	-0.028	(-1.32)
White	0.554	(0.252)	0.504	(0.276)	0.051	(0.95)
Other	0.030	(0.025)	0.055	(0.029)	-0.025***	(-4.71)
Male	0.513	(0.010)	0.513	(0.005)	0.000	(0.11)
Female	0.487	(0.010)	0.487	(0.005)	-0.000	(-0.11)
Students (thousands)	5.655	(9.765)	28.081	(37.502)	-22.426***	(-6.34)
Free and Reduced Lunch	0.668	(0.171)	0.616	(0.146)	0.052	(1.49)
Students with Disabilities	0.123	(0.024)	0.121	(0.018)	0.002	(0.38)
English Language Learner	0.025	(0.037)	0.045	(0.043)	-0.021*	(-2.58)
Retrofits (2007-2017)						
Buses Retrofitted per Retrofit			66.39	(145.3)		
Proportion of Fleet Retrofitted			0.189	(0.141)		
Average Retrofit Cost per Bus ($)			8111.0	(5013.8)		
Bus Fleet Characteristics (2007-2017)						
Average Time in Bus (minutes)	44.883	(11.629)	49.631	(7.940)	-4.748*	(-2.04)
District Bus Ridership	0.621	(0.174)	0.610	(0.087)	0.0113	(0.33)
Total Buses	75.594	(104.432)	313.286	(411.857)	-237.7***	(-6.17)
Total Bus Riders (thousands)	3.475	(7.412)	17.563	(25.814)	-14.088***	(-5.62)
Average Bus Age	14.126	(1.574)	14.268	(1.537)	-0.142	(-0.43)
Observations	153		27		180	

Mean coefficients reported; standard deviations in parentheses. Observations are at the district level. Other demographic category includes Asian, American Indian, Pacific Islander, and Multiracial. Students represents the average student enrollment in thousands. Standardized math and ELA test scores are negative because the majority of Georgia school districts are rural, small, and under-achieving relative to larger urban districts. Aerobic capacity attempts / enrollment represents the number of attempts divided by K-12 enrollment.

Table B.14: FitnessGram Health 2012-2017

	(1) AC	(2) AC	(3) AC	(4) BMI	(5) BMI	(6) BMI
Percent Retrofitted	1.815**			-0.241		
	(0.81)			(0.33)		
Percent Retrofitted * Ridership		2.439*			-0.479	
		(1.36)			(0.53)	
Percent Retrofitted * Ridership * Trip Duration			0.041			-0.010
			(0.03)			(0.01)
Dep. Var. mean	41.66	41.66	41.66	21.03	21.03	21.03
R^2	0.197	0.199	0.198	0.050	0.051	0.051
N	856	846	846	863	853	853

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the district level in parentheses. Year fixed effects included. Demographic variables include the proportion of students that are Asian, African-American, Hispanic, and male, where White and female are the omitted categories, as well as the percentage of students with free or reduced price lunch, disabilities, and English-language learner status. Bus characteristics include average bus age, the proportion of buses built before 2007, and the proportion of liquid natural gas-, butane-, and gasoline-powered buses in the district. Student ridership and trip duration variables also included as controls. The independent variable percent retrofitted is the proportion of a district's bus fleet that is retrofitted in a given year, and zero otherwise. Percent retrofitted * ridership is the percent of the bus fleet retrofitted times the time-constant proportion of students in a district riding the bus, while percent retrofitted * ridership * trip duration is the proportion of the bus fleet retrofitted times time-constant ridership and the time-constant average duration of a daily bus commute for students in a given district.

Table B.15: FitnessGram Health by Gender and School Type 2012-2017

	Elementary		Middle		High School	
	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Percent Retrofitted	3.963**	4.152*	-1.651	0.304	1.899**	1.802
	(1.99)	(2.19)	(1.26)	(2.02)	(0.78)	(1.41)
Dep. Var. mean	42.45	39.82	43.22	38.85	43.45	37.75
R^2	0.143	0.273	0.093	0.305	0.031	0.086
N	777	777	770	770	710	710

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the district level in parentheses. Year fixed effects included. Demographic variables include the proportion of students that are Asian, African-American, Hispanic, and male, where White and female are the omitted categories, as well as the percentage of students with free or reduced price lunch, disabilities, and English-language learner status. Bus characteristics include average bus age, the proportion of buses built before 2007, and the proportion of liquid natural gas-, butane-, and gasoline-powered buses in the district. Student ridership and trip duration variables also included as controls. The independent variable percent retrofitted is the proportion of a district's bus fleet that is retrofitted in a given year, and zero else.

Table B.16: Academic Achievement 2007-2017

	(1) ELA	(2) ELA	(3) ELA	(4) Math	(5) Math	(6) Math	(7) Attend	(8) Attend	(9) Attend
Percent Retrofitted	0.089***			0.049			0.153		
	(0.03)			(0.03)			(0.25)		
Percent Retrofitted * Ridership		0.145***			0.082			0.283	
		(0.04)			(0.05)			(0.40)	
Percent Retrofitted * Ridership * Trip Duration			0.003***			0.001			0.006
			(0.00)			(0.00)			(0.01)
Dep. Var. mean	-0.100	-0.100	-0.100	-0.099	-0.099	-0.099	95.57	95.57	95.57
R^2	0.058	0.058	0.058	0.024	0.024	0.024	0.097	0.097	0.097
N	1,800	1,800	1,800	1,800	1,800	1,800	1,800	1,800	1,800

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the district level in parentheses. Year fixed effects included. Demographic control variables include the proportion of students that are Asian, African-American, Hispanic, and male, where White and female are the omitted categories, as well as the percentage of students with free or reduced price lunch, disabilities, and English-language learner status. Bus control variables include average bus age, the proportion of buses built before 2007, and the proportion of liquid natural gas-, butane-, and gasoline-powered buses in the district. Student ridership and trip duration variables also included as controls. The independent variable percent retrofitted is the proportion of a district's bus fleet that is retrofitted in a given year, and zero else. Percent retrofitted * ridership is the percent of the bus fleet retrofitted times the time-constant proportion of students in a district riding the bus, while percent retrofitted * ridership * trip duration is the proportion of the bus fleet retrofitted times time-constant ridership and the time-constant average duration of a daily bus commute for students in a given district.

Table B.17: Academic Achievement by School Type 2007-2017

	Elementary		Middle	
	ELA	Math	ELA	Math
	(1)	(2)	(3)	(4)
Percent Retrofitted	0.119***	0.061	0.059**	0.047
	(0.03)	(0.07)	(0.03)	(0.03)
Dep. Var. mean	-0.091	-0.089	-0.107	-0.107
R^2	0.043	0.02	0.042	0.037
N	1,800	1,800	1,800	1,800

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the district level in parentheses. Year fixed effects included. Demographic variables include the proportion of students that are Asian, African-American, Hispanic, and male, where White and female are the omitted categories, as well as the percentage of students with free or reduced price lunch, disabilities, and English-language learner status. Bus characteristics include average bus age, the proportion of buses built before 2007, and the proportion of liquid natural gas-, butane-, and gasoline-powered buses in the district. Student ridership and trip duration variables also included as controls. The independent variable percent retrofitted is the proportion of a district's bus fleet that is retrofitted in a given year, and zero else. Elementary includes end-of-grade test scores for grades 3-5, while middle includes the same for grades 6-8.

Table B.18: All Outcomes by Retrofit Type

	(1) ELA	(2) Math	(3) Attend	(4) AC	(5) BMI
Diesel Particulate Filter	0.134**	0.063	0.459	1.411	-0.612
	(0.05)	(0.07)	(0.52)	(1.89)	(0.54)
Closed-Crankcase Filter	-0.022	-0.012	-0.635	-	-
	(0.04)	(0.05)	(0.45)	(.)	(.)
Diesel Oxidation Catalyst	0.051**	0.047	0.144	1.367	-0.139
	(0.02)	(0.03)	(0.19)	(0.85)	(0.46)
Flow-through Filter	-0.026	-0.177***	-0.149	-	-
	(0.06)	(0.05)	(1.43)	(.)	(.)
Dep. Var. mean	-0.100	-0.099	95.57	41.66	21.03
R^2	0.058	0.023	0.096	0.186	0.049
N	1,800	1,800	1,800	856	863

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the district level in parentheses. Year fixed effects included. Demographic control variables include the proportion of students that are Asian, African-American, Hispanic, and male, where White and female are the omitted categories, as well as the percentage of students with free or reduced price lunch, disabilities, and English-language learner status. Student ridership and trip duration variables also included as controls. The number of buses replaced early also included as a control. Bus characteristics not included due to high correlation with covariates. The independent variables each represent the proportion of a bus fleet that is retrofitted with the given engine modification. The sample includes 32 DPF retrofits, nine CCF retrofits, eight DOC retrofits, and three flow-through filter retrofits. Accordingly, results for flow-through filter retrofits may be unreliable.

Table B.19: All Outcomes Fixed Effects Estimates

	(1) ELA	(2) Math	(3) Attend	(4) AC	(5) BMI
I: First Differences					
Percent Retrofitted	0.089***	0.049	0.154	1.815**	-0.241
	(0.03)	(0.03)	(0.25)	(0.81)	(0.33)
R^2	0.058	0.023	0.097	0.197	0.050
N	1,800	1,800	1,800	856	863
II: Fixed Effects					
Percent Retrofitted	0.092	-0.006	0.130	1.108	-0.166
	(0.06)	(0.08)	(0.22)	(1.61)	(0.34)
R^2	0.900	0.900	0.391	0.705	0.613
N	1,958	1,958	1,958	1,034	1,040
D-W F-Stat	197.77	194.42	17.01	31.68	11.39
Prob > F	0.0000	0.0000	0.0001	0.0000	0.0009
III: FE Adding Leads					
Percent Retrofitted	0.100**	0.030	-0.021	0.207	-0.155
	(0.05)	(0.06)	(0.27)	(1.59)	(0.35)
Percent Retrofit Lead	-0.012	-0.047	0.197	5.648*	-0.088
	(0.05)	(0.07)	(0.29)	(3.32)	(0.67)
R^2	0.900	0.900	0.391	0.706	0.613
N	1,958	1,958	1,958	1,034	1,040
IV: FE Adding Leads and Trends					
Percent Retrofitted	0.032	0.037	0.204	1.610*	-0.299
	(0.03)	(0.05)	(0.42)	(0.95)	(0.72)
Percent Retrofit Lead	-0.053	-0.029	0.201	3.091	-0.063
	(0.05)	(0.06)	(0.44)	(4.16)	(1.33)
R^2	0.955	0.943	0.565	0.860	0.756
N	1,958	1,958	1,958	1,034	1,040

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the district level in parentheses. Year and district fixed effects included. Outcomes are district average ELA test scores, Math test scores, attendance, aerobic capacity, and BMI. Demographic variables include the proportion of students that are Asian, African-American, Hispanic, and male, where White and female are the omitted categories, as well as the percentage of students with free or reduced price lunch, disabilities, and English-language learner status. Bus characteristics include average bus age, the proportion of buses built before 2007, and the proportion of liquid natural gas-, butane-, and gasoline-powered buses in the district. Student ridership and trip duration variables also included as controls. The independent variable percent retrofitted is the proportion of a district's bus fleet that has ever been retrofitted, and zero else. The table compares all outcomes using the fixed-effects model to the estimates in our preferred first-differences specification. It also shows how the estimates differ when adding lead treatment variables and district-specific linear trends.

Table B.20: Academic Achievement by Grade 2007-2017

	Grade 3/6		Grade 4/7		Grade 5/8	
	ELA	Math	ELA	Math	ELA	Math
	(1)	(2)	(3)	(4)	(5)	(6)
I. Elementary Schools						
Percent Retrofitted	0.087	0.034	0.208**	0.203*	0.169***	0.060
	(0.07)	(0.11)	(0.10)	(0.11)	(0.05)	(0.08)
R^2	0.033	0.022	0.072	0.064	0.056	0.069
II. Middle Schools						
Percent Retrofitted	0.048	-0.003	0.049	0.031	0.073*	0.108
	(0.05)	(0.06)	(0.05)	(0.04)	(0.04)	(0.08)
R^2	0.048	0.015	0.037	0.024	0.016	0.038
N	1,800	1,800	1,800	1,800	1,800	1,800

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the district level in parentheses. Outcomes are grade-level ELA and math scores. Year fixed effects included. Demographic variables include the proportion of students that are Asian, African-American, Hispanic, and male, where White and female are the omitted categories, as well as the percentage of students with free or reduced price lunch, disabilities, and English-language learner status. Bus characteristics include average bus age, the proportion of buses built before 2007, and the proportion of liquid natural gas-, butane-, and gasoline-powered buses in the district. Student ridership and trip duration variables also included as controls. The independent variables percent retrofitted is the proportion of a district's bus fleet that is retrofitted in a given year, and zero else.

Table B.21: Fixed Effects with Multiple Leads 2007-2017

	(1) ELA	(2) Math	(3) Attend	(4) AC	(5) BMI
Percent Retrofitted	0.031	0.045	0.313	1.610*	-0.299
	(0.04)	(0.05)	(0.43)	(0.95)	(0.72)
Lead 1	-0.050	-0.050	-0.116	3.091	-0.063
	(0.03)	(0.05)	(0.32)	(4.16)	(1.33)
Lead 2	-0.004	0.044	0.660	-	-
	(0.06)	(0.05)	(0.41)	(.)	(.)
R^2	0.955	0.943	0.565	0.860	0.756
N	1,958	1,958	1,958	1,034	1,040

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the district level in parentheses. Year and district fixed effects included. Outcomes are district average ELA test scores, Math test scores, attendance, aerobic capacity, and BMI. Demographic variables include the proportion of students that are Asian, African-American, Hispanic, and male, where White and female are the omitted categories, as well as the percentage of students with free or reduced price lunch, disabilities, and English-language learner status. Bus characteristics include average bus age, the proportion of buses built before 2007, and the proportion of liquid natural gas-, butane-, and gasoline-powered buses in the district. Student ridership and trip duration variables also included as controls. The independent variable percent retrofitted is the cumulative proportion of a district's bus fleet that has ever been retrofitted, and zero else. The table demonstrates the fixed effects estimates when adding two leads to the model. The relevant sample for models (1) and (2) is 2012 - 2017, while model (3) covers the entire sample window, 2007-2017.

Table B.22: Sensitivity of Aerobic Capacity Results to Different Cutoffs 2012-2017

	(1) 15	(2) 20	(3) 25	(4) 30	(5) 35	(6) 30 & 26	(7) Jumps	(8) None	(9) 2012
Percent Retrofitted	2.298	1.313	1.483	1.763**	1.675**	1.815**	3.528***	1.324	7.089***
	(2.07)	(1.34)	(0.97)	(0.73)	(0.70)	(0.81)	(0.85)	(2.29)	(1.09)
R^2	0.248	0.238	0.223	0.218	0.147	0.197	0.098	0.246	0.300
N	860	860	860	857	849	856	675	860	681

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the district level in parentheses. Year fixed effects included. Demographic and bus characteristics included as controls. Student ridership and trip duration variables also included as controls. For models (1) through (6), column headers represent different $V0_2$ max cutoff values. Average aerobic capacity in 2011-12 and 2012-13 is left-skewed, with many implausibly low values for $V0_2Max$. In later years, no school-average $V0_2Max$ is below 30 for male assessments and 26 for female assessments. We therefore demonstrate aerobic capacity results under a range of cutoffs, where each cutoff represents dropping school-level aerobic capacity results below the given value. In column (7), labeled Jumps, we replace as missing any school with average values that increase or decrease by more than 6 $V0_2$ max units from 2011-12 to 2012-13. These jumps are very large in relation to those observed after 2012-13, and so dropping these observations is often equivalent to dropping all values below a given low-valued cutoff. In the column (2012), we drop the entire year of 2011-12, which restricts the number of retrofitting districts such that the coefficient is estimated from only three retrofitting districts. We prefer model (6), the cutoff at 30 for males and 26 for females, because it creates a 2012 distribution that best conforms to the other years of the sample while simultaneously not dropping too many low yet accurate results. In almost all cases, the cutoffs drop less than a tenth of school observations in any given district. Controlling for the proportion of schools dropped does not affect the results because the proportion dropped is not correlated with treatment.

Table B.23: Correlation of Proportion of a Bus Fleet Retrofitted with District Characteristics 2007-2017

	(1) ΔAC Part.	(2) ΔBMI Part.	(3) ΔRidership
I. Endogenous Responses			
Percent Retrofitted	-0.449	-0.576*	-0.023
	(0.32)	(0.30)	(0.04)
R^2	0.153	0.030	0.012
N	870	870	1,780
	Δ Bus Age	Δ Total Buses	Δ Trip Duration
II. Bus Characteristics			
Percent Retrofitted	0.470	56.662	4.227
	(0.63)	(36.02)	(5.80)
R^2	0.129	0.057	0.011
N	1,800	1,800	1,780
	Δ Afr. American	Δ Hispanic	Δ Male
III. Student Demographics			
Percent Retrofitted	0.015	0.466	-0.109
	(0.36)	(0.38)	(0.30)
R^2	0.021	0.028	0.007
N	1,800	1,800	1,800
	Δ ELL	Δ SWD	Δ FRPL
IV. Student Characteristics			
Percent Retrofitted	-0.009	0.572**	0.924
	(0.08)	(0.29)	(1.58)
R^2	0.012	0.219	0.035
N	1,800	1,800	1,800

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the district level in parentheses. Year fixed effects included. In the first panel, models (1) and (2) demonstrate the extent to which the proportion of a bus fleet retrofitted is correlated with changes in the participation rate, i.e., the number of attempts divided by district enrollment. Model (3) shows whether the proportion of a bus fleet retrofitted is correlated with year-on-year changes the ridership rate. The relevant sample for models (1) and (2) is 2012 - 2017, while model (3) covers the entire sample window, 2007-2017. In the second panel, we show that the proportion of a bus fleet retrofitted is not significantly correlated with changes in the average bus age within a district, the total number of buses, or the average trip duration. The third panel demonstrates that the proportion of a bus fleet retrofitted is not significantly related to changes in the percent of a district that is African American, Hispanic, or Male. The fourth panel shows the relationship between the proportion of a bus fleet retrofitted and changes in the percent of a district's students that are English language learner, students with disabilities, or receiving free- and reduced-price lunch.

Table B.24: Academic Achievement 2007-2017, Dropping Milestones Years 2015-2017

	(1) ELA	(2) ELA	(3) ELA	(4) Math	(5) Math	(6) Math
Percent Retrofitted	0.089***			0.049		
	(0.03)			(0.03)		
Percent Retrofitted * Ridership		0.143***			0.083*	
		(0.04)			(0.05)	
Percent Retrofitted * Ridership * Trip Duration			0.002***			0.001
			(0.00)			(0.00)
R^2	0.064	0.064	0.064	0.020	0.020	0.020
N	1,440	1,440	1,440	1,440	1,440	1,440

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the district level in parentheses. Outcomes are district average ELA test scores, Math test scores, attendance, aerobic capacity, and BMI. District and year fixed effects included. Demographic variables include the proportion of students that are Asian, African-American, Hispanic, and male, where White and female are the omitted categories, as well as the percentage of students with free or reduced price lunch, disabilities, and English-language learner status. Bus characteristics include average bus age, the proportion of buses built before 2007, and the proportion of liquid natural gas-, butane-, and gasoline-powered buses in the district. Controls for ridership share and trip duration are also included. The table shows how our first-differences estimates change when dropping all years after 2014-15 when the new Milestones standardized examination is offered instead of the CRCT exam. Milestones computerized examinations suffered from widespread glitches that may have affected our estimates.

Table B.25: Drop Interpolated Bus Years

	(1) ELA	(2) Math	(3) Attend	(4) AC	(5) BMI
Percent Retrofitted	0.083***	0.057	0.242	1.766**	-0.242
	(0.03)	(0.04)	(0.29)	(0.83)	(0.35)
R^2	0.079	0.029	0.174	0.188	0.061
N	1,260	1,260	1,260	692	698

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the district level in parentheses. Outcomes are district average ELA test scores, Math test scores, attendance, aerobic capacity, and BMI. District and year fixed effects included. Demographic variables include the proportion of students that are Asian, African-American, Hispanic, and male, where White and female are the omitted categories, as well as the percentage of students with free or reduced price lunch, disabilities, and English-language learner status. Bus characteristics include average bus age, the proportion of buses built before 2007, and the proportion of liquid natural gas-, butane-, and gasoline-powered buses in the district. Controls for ridership share and trip duration are also included. The independent variable percent retrofitted is the proportion of a district's bus fleet that is retrofitted in a given year, and zero else. The table shows how our first-differences estimates change when dropping all years for which information on district bus fleets is lacking. For these years, we inserted the value of the nearest year for which data is available, which is 2010 for all years prior and 2016 for 2017.

Table B.26: Timing of Retrofit Implementation

	(1) ELA	(2) Math	(3) Attend	(4) AC	(5) BMI
I. Regular Timing					
Percent Retrofitted	0.089***	0.049	0.154	1.815**	-0.241
	(0.03)	(0.03)	(0.25)	(0.81)	(0.33)
R^2	0.058	0.023	0.097	0.197	0.050
N	1,800	1,800	1,800	856	863
II. Treatment 1-year in Advance					
Percent Retrofitted	-0.029	-0.040	-0.110	2.258	0.197
	(0.03)	(0.04)	(0.23)	(2.75)	(0.62)
R^2	0.056	0.023	0.097	0.196	0.050
N	1,800	1,800	1,800	856	863
III. January Implementation					
Percent Retrofitted	0.100***	0.079**	0.273	1.664*	-0.089
	(0.02)	(0.03)	(0.26)	(0.87)	(0.28)
R^2	0.059	0.024	0.097	0.197	0.050
N	1,800	1,800	1,800	856	863

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors in parentheses. Year fixed effects included. Demographic control variables include the proportion of students that are Asian, African-American, Hispanic, and male, where White and female are the omitted categories. The percentage of students with free or reduced price lunch, disabilities, and English-language learner status. Bus characteristics include average bus age, the proportion of buses built before 2007, and the proportion of liquid natural gas-, butane-, and gasoline-powered buses in the district. Student ridership and trip duration variables also included as controls. The independent variable percent retrofitted varies according to the timing of the retrofit completion date in a given school district. In the base case presented in Panel I, regular timing, all retrofits between May and the following April are assigned to the fiscal year of the latter April. In Panel II, we show the results when assigning a placebo treatment year as one year before the actual retrofit completion year. In Panel III, retrofits completed before January are assigned to the same fiscal year, but those occurring after January are assigned to the following fiscal year.

Table B.27: FitnessGram Health with Linear Trends 2012-2017

	(1) AC	(2) AC	(3) AC	(4) BMI	(5) BMI	(6) BMI
Percent Retrofitted	2.197***			-0.345		
	(0.60)			(0.56)		
Percent Retrofitted * Ridership		3.196***			-0.660	
		(1.05)			(0.88)	
Percent Retrofitted * Ridership * Trip Duration			0.055**			-0.016
			(0.02)			(0.01)
Dep. Var. mean	41.66	41.66	41.66	21.03	21.03	21.03
R^2	0.353	0.353	0.353	0.200	0.200	0.200
N	846	846	846	853	853	853

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the district level in parentheses. Year fixed effects and district-fixed effects, to account for linear trends, included. Demographic variables include the proportion of students that are Asian, African-American, Hispanic, and male, where White and female are the omitted categories, as well as the percentage of students with free or reduced price lunch, disabilities, and English-language learner status. Bus characteristics include average bus age, the proportion of buses built before 2007, and the proportion of liquid natural gas-, butane-, and gasoline-powered buses in the district. Student ridership and trip duration variables also included as controls. The independent variable percent retrofitted is the proportion of a district's bus fleet that is retrofitted in a given year, and zero otherwise. Percent retrofitted * ridership is the percent of the bus fleet retrofitted times the time-constant proportion of students in a district riding the bus, while percent retrofitted * ridership * trip duration is the proportion of the bus fleet retrofitted times time-constant ridership and the time-constant average duration of a daily bus commute for students in a given district.

Table B.28: Academic Achievement with Linear Trends 2007-2017

	(1) ELA	(2) ELA	(3) ELA	(4) Math	(5) Math	(6) Math	(7) Attend	(8) Attend	(9) Attend
Percent Retrofitted	0.076***			0.045			0.107		
	(0.03)			(0.04)			(0.35)		
Percent Retrofitted		0.121**			0.075			0.195	
* Ridership		(0.05)			(0.06)			(0.56)	
Percent Retrofitted			0.002**			0.001			0.004
* Ridership * Trip Duration			(0.00)			(0.00)			(0.01)
Dep. Var. mean	-0.100	-0.100	-0.100	-0.099	-0.099	-0.099	95.57	95.57	95.57
R^2	0.137	0.137	0.137	0.070	0.070	0.070	0.153	0.153	0.153
N	1,780	1,780	1,780	1,780	1,780	1,780	1,780	1,780	1,780

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the district level in parentheses. Year fixed effects and district-fixed effect, to account for linear trends, included. Demographic control variables include the proportion of students that are Asian, African-American, Hispanic, and male, where White and female are the omitted categories, as well as the percentage of students with free or reduced price lunch, disabilities, and English-language learner status. Bus control variables include average bus age, the proportion of buses built before 2007, and the proportion of liquid natural gas-, butane-, and gasoline-powered buses in the district. Student ridership and trip duration variables also included as controls. The independent variable percent retrofitted is the proportion of a district's bus fleet that is retrofitted in a given year, and zero else. Percent retrofitted * ridership is the percent of the bus fleet retrofitted times the time-constant proportion of students in a district riding the bus, while percent retrofitted * ridership * trip duration is the proportion of the bus fleet retrofitted times time-constant ridership and the time-constant average duration of a daily bus commute for students in a given district.

Table B.29: School District and County Characteristics by Farm-to-School Policy Status

	(1) Farm-to-Schools		(2) Not Farm-to-Schools		(3) T-test of Means
District Characteristics (2001-2017)					
Total Enrollment	325.9	(486.9)	66.10	(69.53)	-259.8***
FRL Share	0.614	(0.139)	0.699	(0.133)	0.0842***
Nutrition Expenditures	9423.6	(13207.0)	2140.5	(2078.6)	-7283.1***
% CEP Schools	0.434	(0.452)	0.572	(0.444)	0.138
Marginal CEP Students	521.6	(865.3)	271.5	(480.6)	-250.1*
% Marginal CEP Students	0.0331	(0.0496)	0.0448	(0.0666)	0.0117
Nutritional Expenditures (2001-2017)					
Nutrition Expenditures	9423.6	(13207.0)	2140.5	(2078.6)	-7283.1***
Federal Nutrition Revenue	7813.5	(10945.6)	1743.7	(1679.9)	-6069.8***
Local Nutrition Revenue	2109.8	(4248.2)	349.6	(587.4)	-1760.3***
State Nutrition Revenue	261.4	(381.6)	57.28	(59.10)	-204.1***
County Agricultural Revenues (2000-2018)					
All Revenues	62078.5	(78315.1)	61559.4	(71870.8)	-519.1
Animal Products	43250.6	(77299.2)	36796.5	(60137.3)	-6454.1
Fruits and Vegetables	15022.4	(20775.5)	20596.7	(28615.2)	5574.2
Agrotourism	15.41	(39.91)	8.340	(34.19)	-7.067
Meats	37124.2	(66703.6)	31504.8	(53829.2)	-5619.4
Dairy	1412.4	(4093.8)	2317.3	(6096.7)	904.9
School Visits	15.41	(39.91)	8.340	(34.19)	-7.067
Other Revenues	4338.6	(5287.1)	6664.8	(7485.0)	2326.2*
Placebo Revenues	6477.6	(8333.4)	13574.6	(12885.0)	7097.0***
Farm-to-School Policy Variables (2001-2017)					
Farm-to-School Enrollment	107.1	(212.0)			
Marginal CEP Students in FtS District	443.1	(720.8)			
% Marginal CEP Students in FtS District	0.0264	(0.0392)			
District Observations	75		105		180
County Observations	63		96		159

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Mean coefficients reported in columns (1) and (2); standard deviations in parentheses. All dollar amounts expressed in thousands of 2017 dollars. Table 5.35 lists the commodities included in each category.

Table B.30: Agricultural Revenues 2001-2017

	OLS		Spatial Lag 1		Spatial Lag 2		CEP 1		CEP 2	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Within-County Effects										
Nutrition	0.174	0.147	0.180	0.152	0.174	0.148	0.199	0.165	0.158	0.132
Expenditures	(0.386)	(0.392)	(0.404)	(0.404)	(0.406)	(0.407)	(0.404)	(0.404)	(0.408)	(0.408)
FtS Policy		1913.4		1927.2		2178.1		3365.0*		3282.3
		(5072.9)		(1894.4)		(1903.2)		(2038.0)		(2074.4)
CEP							25167.5	43072.4**	25984.1	46586.1**
							(19544.0)	(21809.2)	(19471.9)	(21911.4)
CEP × FtS								-75222.3*		-76044.2*
Policy								(43600.1)		(43686.6)
Contiguous County Effects										
Nutrition					-0.943	-0.831			-1.239	-1.168
Expenditures					(0.907)	(0.911)			(0.914)	(0.916)
FtS Policy						-6537.3				-8274.6
						(4854.7)				(5172.3)
CEP									85222.5*	95942.9*
									(47895.0)	(53648.6)
CEP × FtS										-39083.0
Policy										(114514.2)
Implied Δ Revenue		550M		555M		627M	390M	683M	402M	656M
N	2,669	2,669	2,669	2,669	2,669	2,669	2,669	2,669	2,669	2,669

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Year and county fixed effects included in all models. Time-varying control variables include the total number of students in each district and FRL shares. Clustered standard errors at the county level in models (1) and (2). Standard errors clustered across contiguous counties in models (3) - (10). Models (3) - (10) include dependent variable lags across inverse distance bandwidths, controlling for agricultural revenues in county i by agricultural revenues in all counties j in proportion to the inverse distance between county i and county j. $FtSPolicy$ is the proportion of students in a county served by a farm-to-school policy in their school district. CEP is the proportion of students in a county that are induced into free lunch by their schools participation in the Community Eligibility Provision of the HHFKA. $CEP * FtS$ is the proportion of students in a county that are induced into free lunch by their schools participation in the Community Eligibility Provision of the HHFKA while also being served by farm-to-school districts. Coefficients on nutrition expenditures are scaled to the dollar, while coefficients on FtS Policy, CEP, and CEP*FtS are scaled to the each thousand dollars. Implied Δ Revenue is the estimated change in statewide agricultural revenues associated with the policy coefficient(s) of interest in each model. In models (1) through (6), this is simply equal to the point estimate times in row (2) times the total number of farm-to-school policy county-years (i.e., 288). In models (7) and (9), implied change in revenue is the point estimate in row (3) times the sum of CEP shares (i.e., 15.5). In models (8) and (10), the implied revenue change is the point estimate in row (2) times 288 subtracted by the point estimate in (4) times the sum of all CEP*FtS combinations (i.e., 3.8).

Table B.31: Agricultural Revenues - Animal Products 2001-2017

	OLS		Spatial Lag 1	Spatial Lag 2	CEP 1	CEP 2
	(1)	(2)	(3)	(4)	(5)	(6)
Within-County Effects						
Nutrition Expenditures	0.229	0.165	0.193	0.160	0.199	0.0868
	(0.285)	(0.287)	(0.333)	(0.332)	(0.333)	(0.333)
FtS Policy		4531.3	4630.8***	4587.7***	5175.6***	4312.6**
		(4844.7)	(1553.7)	(1552.3)	(1678.4)	(1693.7)
CEP					12130.3	11682.8
					(18319.4)	(17957.1)
CEP × FtS Policy					-29202.4	-26095.5
					(35897.5)	(35860.8)
Contiguous County Effects						
Nutrition Expenditures				0.314		-0.0336
				(0.781)		(0.780)
FtS Policy				362.7		695.6
				(4226.3)		(4497.0)
CEP						155718.3***
						(45202.6)
CEP × FtS Policy						-168525.1*
						(96327.1)
N	2,669	2,669	2,669	2,669	2,669	2,669

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Year and county fixed effects included in all models. Time-varying control variables include the total number of students in each district and FRL shares. Clustered standard errors at the county level in models (1) and (2). Standard errors clustered across contiguous counties in models (3) - (6). Models (3) - (6) include dependent variable lags across inverse distance bandwidths, controlling for agricultural revenues in county i by agricultural revenues in all counties j in proportion to the inverse distance between county i and county j. $FtSPolicy$ is the proportion of students in a county served by a farm-to-school policy in their school district. CEP is the proportion of students in a county that are induced into free lunch by their schools participation in the Community Eligibility Provision of the HHFKA. $CEP * FtS$ is the proportion of students in a county that are induced into free lunch by their schools participation in the Community Eligibility Provision of the HHFKA while also being served by farm-to-school districts. Coefficients on nutrition expenditures are scaled to the dollar, while coefficients on FtS Policy, CEP, and CEP*FtS are scaled to the each thousand dollars.

Table B.32: Agricultural Revenues - Fruits and Vegetables 2001-2017

	OLS		Spatial Lag 1	Spatial Lag 2	CEP 1	CEP 2
	(1)	(2)	(3)	(4)	(5)	(6)
Within-County Effects						
Nutrition	0.0119	0.0486	0.113	-0.0399	0.137	-0.0120
Expenditures	(0.314)	(0.301)	(0.178)	(0.197)	(0.178)	(0.197)
FtS Policy		-2575.9	-1858.9**	-2416.0***	-891.6	-1645.2*
		(1689.1)	(839.1)	(879.7)	(910.8)	(957.3)
CEP					24775.8***	24919.9**
					(9563.0)	(10286.5)
CEP × FtS					-44989.4**	-32876.3
Policy					(19067.2)	(21109.3)
Contiguous County Effects						
Nutrition				-0.920*		-0.966*
Expenditures				(0.000536)		(0.000537)
FtS Policy				-5820.6**		-8154.6***
				(2592.6)		(2863.0)
CEP						-16419.4
						(30769.7)
CEP × FtS						94053.6
Policy						(65743.7)
N	2,669	2,669	2,669	2,669	2,669	2,669

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Year and county fixed effects included in all models. Time-varying control variables include the total number of students in each district and FRL shares. Clustered standard errors at the county level in models (1) and (2). Standard errors clustered across contiguous counties in models (3) - (6). Models (3) - (6) include dependent variable lags across inverse distance bandwidths, controlling for agricultural revenues in county i by agricultural revenues in all counties j in proportion to the inverse distance between county i and county j. $FtSPolicy$ is the proportion of students in a county served by a farm-to-school policy in their school district. CEP is the proportion of students in a county that are induced into free lunch by their schools participation in the Community Eligibility Provision of the HHFKA. $CEP*FtS$ is the proportion of students in a county that are induced into free lunch by their schools participation in the Community Eligibility Provision of the HHFKA while also being served by farm-to-school districts. Coefficients on nutrition expenditures are scaled to the dollar, while coefficients on FtS Policy, CEP, and CEP*FtS are scaled to the each thousand dollars.

Table B.33: Agricultural Revenues - Agrotourism 2001-2017

	OLS		Spatial Lag 1	Spatial Lag 2	CEP 1	CEP 2
	(1)	(2)	(3)	(4)	(5)	(6)
Within-County Effects						
Nutrition Expenditures	-0.0435	-0.0441	-0.0431	-0.0420	-0.0469*	-0.0447*
	(0.0337)	(0.0339)	(0.0264)	(0.0264)	(0.0263)	(0.0264)
FtS Policy		46.83	51.98	54.41	-83.06	-73.72
		(117.1)	(123.7)	(123.4)	(132.8)	(134.2)
CEP					-6259.7***	-6101.8***
					(1411.5)	(1421.9)
CEP × FtS Policy					5960.9**	6061.6**
					(2837.8)	(2856.7)
Contiguous County Effects						
Nutrition Expenditures				-0.0209		-0.00485
				(0.0617)		(0.0615)
FtS Policy				-199.4		-259.0
				(326.2)		(345.9)
CEP						-1118.1
						(3805.4)
CEP × FtS Policy						11489.3
						(7836.7)
N	2,669	2,669	2,669	2,669	2,669	2,669

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Year and county fixed effects included in all models. Time-varying control variables include the total number of students in each district and FRL shares. Clustered standard errors at the county level in models (1) and (2). Standard errors clustered across contiguous counties in models (3) - (6). Models (3) - (6) include dependent variable lags across inverse distance bandwidths, controlling for agricultural revenues in county i by agricultural revenues in all counties j in proportion to the inverse distance between county i and county j. $FtSPolicy$ is the proportion of students in a county served by a farm-to-school policy in their school district. CEP is the proportion of students in a county that are induced into free lunch by their schools participation in the Community Eligibility Provision of the HHFKA. $CEP * FtS$ is the proportion of students in a county that are induced into free lunch by their schools participation in the Community Eligibility Provision of the HHFKA while also being served by farm-to-school districts. Coefficients on nutrition expenditures are scaled to the dollar, while coefficients on FtS Policy, CEP, and CEP*FtS are scaled to the each thousand dollars.

Table B.34: Two Stage Least Squares: Agricultural Revenues 2001-2017

	(1) Total Revenues	(2) Animal Products	(3) Fruits and Vegetables
First Stage			
CEP Share	2,696,274***	2,696,274***	2,696,274***
	(290860.1)	(290860.1)	(290860.1)
F-Stat of Excluded Instruments	17.01	17.01	17.01
Second Stage			
Nutrition Expenditures	26.37*	3.86	19.26*
	(15.79)	(5.59)	(10.88)
County Mean Agricultural Revenue	$61.8M	$39.4M	$18.44M
County Mean Nutrition Expenditure	$5.13M	$5.13M	$5.13M
Implied Average County Revenue Change	$354,560	$52,038	$259,651
Implied Local Expenditure Share	6.9%	1.0%	5.0%
Implied Total Revenue Change	$966M	$138M	$690M
Trend Controls	✓	✓	✓
Year & District FEs	✓	✓	✓
Observations	2,701	2,701	2,701

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Clustered standard errors at the county level. Controls include county-specific trends and year and district fixed effects. County mean agricultural revenue is the revenue across all relevant agricultural products in a county, while mean nutrition expenditures is the total nutrition expenditures by schools in a county. Implied Average County Revenue Change is the first-stage CEP coefficient times the average CEP share and the second stage coefficient, representing the increase in local agricultural revenues associated with increasing numbers of free lunch students through with the CEP across each county-year combination on average. Implied local expenditure share is the average proportion of nutrition expenditures that appear to be recouped by farmers in the same county. Implied Total Revenue Change is the total nutritional outlays over the entire sample ($13.8B) times the implied local expenditure share.

Table B.35: Commodities Included in Each Agricultural Category

Main Categories	Commodities
Animal Products	Beef, Catfish, Chicken, Dairy, Eggs, Fishing, Pork
Fruits & Vegetables	Apples, Banana Peppers, Barley, Bell Peppers, Blackberries, Blueberries, Broccoli, Cabbage, Cantaloupe, Carrots, Collards, Container Nursery, Corn, Cucumbers, Eggplant, English Peas, Field Nursery, Green House, Grapes, Green Onions, Hay, Honey Bees, Hot Peppers, Irish Potatoes, Kale, Lettuce, Lima Beans, Mustard, Organics, Oats, Okra, Onions, Peaches, Peanuts, Pecans, Pole Beans, Pumpkin, Rye, Snap Beans, Sorghum, Southern Peas, Soybeans, Spinach, Strawberries, Sweet Corn, Tomato, Turnip Greens, Turnip Roots, Watermelon, Winter Squash, Yellow Squash, Zucchini
Agrotourism	Corn Maze, Guide Services, Hayrides, School tours, Special Attractions, Special Events
Other	Other , Silage, Pine Straw, Straw, Turfgrass
Placebo Commodities	Timber, Camping, Christmas Trees, Cotton, Hunting Leases, Tobacco
Unused Commodities	Horses, Wildlife Observation, Goats, Quail, Sheep

Table B.36: Golden Radish Application Information 2014-2018

	Mean	SD	% Missing
Days Local	118.2	(67.97)	70.55
Local Meals	1110443.7	(1890232.2)	70.74
Taste Tests	134.1	(590.8)	72.09
Farmer Field Trips	13.75	(21.75)	72.75
Local Food Promotions	112.0	(321.1)	71.32
Local Food Lessons	97.21	(408.3)	73.41
Schools with Gardens	12.73	(19.51)	71.65
Food Activities	50.99	(138.3)	73.74
Activities with Committee Members	24.27	(51.85)	73.19
Professional Development Trainings	16.94	(110.7)	73.41
Golden Radish Awards	197	-	
Year-District Observations	910		

Mean coefficients reported; standard deviations in parentheses.

Table B.37: Community Eligibility Provision Predicts Nutrition Expenditures 2001-2017

	(1)	(2)	(3)	(4)	(5)
CEP Students	80.88**				
	(2.18)				
Marginal CEP Students		1385.4**			
		(2.17)			
Marginal FtS CEP Students			1989.7*		
			(1.84)		
% Marginal CEP Students				3068731.1***	
				(4.37)	
% Marginal FtS CEP Students					7427600.5***
					(2.91)
Implied Δ Expenditures	8.3M	77.5M	49.7M	36M	20.7M
R2	0.704	0.699	0.700	0.693	0.693
N	2,669	2,669	2,669	2,669	2,669

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. T-statistics in parentheses. Year and county fixed effects included in all models. Clustered standard errors at the county level. Time-varying control variables include the total number of students in each district. $CEPStudents$ is the total number of students at CEP-participating schools in county i in year t. $MarginalCEPStudents$ is the number of students induced into free lunch status by CEP participation in county i in year t. $MarginalFtSCEPStudents$ is the number of students induced into free lunch status by CEP participation in county i in year t, where county i has school districts that participate in the farm-to-school program. $\%MarginalCEPStudents$ is the share of students induced into free lunch status in county i in year t. $\%MarginalFtSCEPStudents$ is the share of students induced into free lunch status in farm-to-school-participating schools in county i in year t. Implied Δ Expenditures is the estimated change in county nutrition expenditures associated with the coefficients of interest in each model, where the implied change in expenditures is calculated by multiplying the point estimate times the all-sample total of the dependent variables. For example, 8.3M in column (1) was calculated by multiplying 80 times the total number of students in CEP participating schools in the sample, which is 102,734.

Appendix C

Bibliography

Bibliography

W. Adamowicz, M. Dickie, S. Gerking, M. Veronesi, and D. Zinner. Household decision making and valuation of environmental health risks to parents and their children. *Journal of the Association of Environmental and Resource Economists*, 1(4):481–519, 2014.

S. D. Adar, J. D'Souza, L. Sheppard, J. D. Kaufman, T. S. Hallstrand, M. E. Davey, J. R. Sullivan, J. Jahnke, J. Koenig, T. V. Larson, et al. Adopting clean fuels and technologies on school buses. pollution and health impacts in children. *American journal of respiratory and critical care medicine*, 191(12):1413–1421, 2015.

D. Alexander and H. Schwandt. The impact of car pollution on infant and child health: Evidence from emissions cheating. IZA Discussion Papers 12427, Institute of Labor Economics (IZA), 2019. URL https://EconPapers.repec.org/RePEc:iza:izadps:dp12427.

D. Almond and J. Currie. Killing me softly: The fetal origins hypothesis. *The Journal of Economic Perspectives*, 25(3):153–172, 2011. ISSN 08953309. URL http://www.jstor.org/stable/23049427.

J.-P. Amigues, M. Moreaux, and K. Schubert. Optimal use of a polluting non-renewable resource generating both manageable and catastrophic damages. *Annals of Economics and Statistics*, (103/104):107–141, 2011. ISSN 21154430, 19683863. URL http://www.jstor.org/stable/41615496.

M. L. Anderson, J. Gallagher, and E. R. Ritchie. School meal quality and academic performance. *Journal of Public Economics*, 168:81–93, 2018.

S. C. Ash. Gis shapefile of all impoundments. http://www.southeastcoalash.org/3, 2019.

A. Baba and A. Kaya. Leaching characteristics of solid wastes from thermal power plants of western turkey and comparison of toxicity methodologies. *Journal of environmental management*, 73:199–207, 12 2004. doi: 10.1016/j.jenvman.2004.06.005.

H. S. Banzhaf and R. P. Walsh. Do people vote with their feet? an empirical test of tiebout. *American Economic Review*, 98(3):843–63, June 2008. doi: 10.1257/aer.98.3.843. URL http://www.aeaweb.org/articles?id=10.1257/aer.98.3.843.

T. L. Barone, J. M. Storey, and N. Domingo. An analysis of field-aged diesel particulate filter performance: Particle emissions before, during, and after regeneration. *Journal of the Air & Waste Management Association*, 60(8):968–976, 2010. doi: 10.3155/1047-3289.60.8.968. URL https://doi.org/10.3155/1047-3289.60.8.968.

A. Barreca, K. Clay, and J. Tarr. Coal, smoke, and death: Bituminous coal and american home heating. Working Paper 19881, National Bureau of Economic Research, February 2014. URL http://www.nber.org/papers/w19881.

T. Beatty and J. Shimshack. School buses, diesel emissions, and respiratory health. *Journal of Health Economics*, 30(5):987–999, 2011. URL http://EconPapers.repec.org/RePEc:eee:jhecon:v:30:y:2011:i:5:p:987-999.

T. K. Beatty and J. P. Shimshack. Air pollution and children's respiratory health: A cohort analysis. *Journal of Environmental Economics and Management*, 67(1):39–57, 2014.

K. Belitz, B. Jurgens, and T. Johnson. Langelier saturation indices computed for u.s. groundwater, 1991-2015; water well data and characteristic values for states: U.s. geological survey data release. *USGS*, 2016. doi: http://dx.doi.org/10.5066/F7XW4GWX. URL https://www.sciencebase.gov/catalog/item/56f30527e4b0f59b85df12fc.

L. S. Bennear and S. M. Olmstead. The impacts of the “right to know”: Information disclosure and the violation of drinking water standards. *Journal of Environmental Economics and Management*, 56(2):117 – 130, 2008. ISSN 0095-0696. doi: https://doi.org/10.1016/j.jeem.2008.03.002. URL http://www.sciencedirect.com/science/article/pii/S0095069608000399.

L. S. Bennear, K. K. Jessoe, and S. M. Olmstead. Sampling out: regulatory avoidance and the total coliform rule. *Environmental science technology*, 43 14:5176–82, 2009.

C. Bernhardt, A. Russ, and E. Schaeffer. Toxic wastewater from coal plants. Technical report, Environmental Integrity Project, 8 2016. URL https://www.eenews.net/assets/2016/08/11/document_gw_05.pdf.

S. Biswas, V. Verma, J. J. Schauer, and C. Sioutas. Chemical speciation of pm emissions from heavy-duty diesel vehicles equipped with diesel particulate filter (dpf) and selective catalytic reduction (scr) retrofits. *Atmospheric Environment*, 43(11):1917 – 1925, 2009. ISSN 1352-2310. doi: https://doi.org/10.1016/j.atmosenv.2008.12.040. URL http://www.sciencedirect.com/science/article/pii/S1352231008011862.

C. Bjornson and I. Mitchell. Gender differences in asthma in childhood and adolescence. *Journal of Gender Specific Medicine*, 3(8):57 – 61, 2000. URL https://www.ncbi.nlm.nih.gov/pubmed/11253268.

K. Blasingame. Measurement agreement of fitnessgram aerobic capacity and body composition standards. *Iowa State University Graduate Theses and Dissertations*, 2012. doi: https://doi.org/10.31274/etd-180810-747. URL https://lib.dr.iastate.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=3288&context=etd.

J. Borak and G. Sirianni. Studies of self-pollution in diesel school buses: methodological issues. *Journal of occupational and environmental hygiene*, 4(9):660–668, 2007.

E. R. Botkins and B. E. Roe. Understanding participation in farm to school programs: Results integrating school and supply-side factors. *Food Policy*, 74:126 – 137, 2018. ISSN 0306-

9192. doi: https://doi.org/10.1016/j.foodpol.2017.12.006. URL http://www.sciencedirect.com/science/article/pii/S0306919217310230.

J. Boyce and M. Ash. Toxic 100 water polluters index: 2016 report, based on 2014 data. Technical Report 2, Political Economy Research Institute, Office of Water, Washington DC, 20460, 9 2016. URL http://www.peri.umass.edu/publication/item/765-toxic-100-names-top-climate-polluters.

E. Brainerd and N. Menon. Seasonal effects of water quality: The hidden costs of the green revolution to infant and child health in india. *Journal of Development Economics*, 107(C):49–64, 2014. URL https://EconPapers.repec.org/RePEc:eee:deveco:v:107:y:2014:i:c:p:49-64.

L. Calderón-Garcidueñas, A. Mora-Tiscareño, M. Styner, G. Gómez-garza, H. Zhu, R. Torres-Jardón, E. Carlos, E. Solorio-López, H. Medina-Cortina, M. Kavanaugh, and A. D'Angiulli. White matter hyperintensities, systemic inflammation, brain growth, and cognitive functions in children exposed to air pollution. *Journal of Alzheimer's Disease*, 31(1):183–191, 2012. URL https://content.iospress.com/articles/journal-of-alzheimers-disease/jad120610.

K. P. Cantor, C. M. Villanueva, D. Silverman, J. D. Figueroa, F. Real, M. Garcia-Closas, N. Malats, S. Chanock, M. Yeager, A. Tardon, R. Garcia-Closas, C. Serra, A. Carrato, G. Castaño-Vinyals, C. Samanic, N. Rothman, and M. Kogevinas. Polymorphisms in gstt1, gstz1, and cyp2e1, disinfection by-products, and risk of bladder cancer in spain. *Environmental Health Perspectives*, 118(11):1545–1550, 11 2010. ISSN 0091-6765. doi: 10.1289/ehp.1002206.

D. M. Castelli, C. H. Hillman, S. M. Buck, and H. E. Erwin. Physical fitness and academic achievement in third-and fifth-grade students. *Journal of Sport and Exercise Psychology*, 29(2): 239–252, 2007.

M. J. Castillo-Garzón, J. Ruiz, F. B. Ortega, and A. Gutiérrez. Anti-aging therapy through fitness enhancement. 1:213–20, 02 2006.

J. Castro-Piñero, A. Perez-Bey, V. Segura-Jiménez, V. A. Aparicio, S. Gómez-Martínez, R. Izquierdo-Gomez, A. Marcos, J. R. Ruiz, A. Marcos, J. Castro-Piñero, et al. Cardiorespiratory fitness cutoff points for early detection of present and future cardiovascular risk in children: a 2-year follow-up study. In *Mayo Clinic Proceedings*, volume 92, pages 1753–1762. Elsevier, 2017.

K. Y. Chay and M. Greenstone. The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. *The Quarterly Journal of Economics*, 118(3):1121–1167, 2003. ISSN 00335533, 15314650. URL http://www.jstor.org/stable/25053932.

X. Chen, X. Zhang, and X. Zhang. Smog in Our Brains: Gender Differences in the Impact of Exposure to Air Pollution on Cognitive Performance. *GLO Discussion Paper Series*, (32), 2017. URL https://ideas.repec.org/p/zbw/glodps/32.html.

R. Chetty, J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan. How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics*, 126(4):1593–1660, 2011. URL https://ideas.repec.org/a/oup/qjecon/v126y2011i4p1593-1660.html.

L. Christensen, B. Jablonski, L. Stephens, and A. Joshi. Evaluating the economic impacts of farm-to-school procurement. *Journal of Agriculture, Food Systems, and Community Development*, 8(C):73–94, Dec. 2018. doi: 10.5304/jafscd.2019.08C.002. URL https://www.foodsystemsjournal.org/index.php/fsj/article/view/656.

L. O. Christensen, B. B. R. Jablonski, and J. K. O'Hara. School districts and their local food supply chains. *Renewable Agriculture and Food Systems*, 34(3):207–215, 2019a. doi: 10.1017/S1742170517000540.

P. Christensen, D. Keiser, and G. Lade. Economic effects of environmental crises: Evidence from flint, michigan. *Working Paper*, 2019b. URL https://drive.google.com/file/d/19PqBMoAfJ-QqGW-Ar6Yae5Ls-jGSFcgV/view.

K. Clay, W. Troesken, and M. R. Haines. Lead and mortality. Working Paper 16480, National Bureau of Economic Research, October 2010. URL http://www.nber.org/papers/w16480.

K. Clay, J. Lewis, and E. Severnini. Pollution, infectious disease, and mortality: Evidence from the 1918 spanish influenza pandemic. Working Paper 21635, National Bureau of Economic Research, October 2015. URL http://www.nber.org/papers/w21635.

K. Clay, J. Lewis, and E. Severnini. Canary in a coal mine: Infant mortality, property values, and tradeoffs associated with mid-20th century air pollution. Working Paper 22155, National Bureau of Economic Research, April 2016. URL http://www.nber.org/papers/w22155.

K. Clay, M. Portnykh, and E. Severnini. Toxic truth: Lead and fertility. Working Paper 24607, National Bureau of Economic Research, May 2018. URL http://www.nber.org/papers/w24607.

K. Clay, M. Portnykh, and E. Severnini. The legacy lead deposition in soils and its impact on cognitive function in preschool-aged children in the united states. *Economics Human Biology*, 33:181 – 192, 2019. ISSN 1570-677X. doi: https://doi.org/10.1016/j.ehb.2019.03.001. URL http://www.sciencedirect.com/science/article/pii/S1570677X18303174.

J. E. Clougherty and L. D. Kubzansky. Traffic-related air pollution and stress: Effects on asthma. *Environmental Health Perspectives*, 116(9):A376–A377, 2008. ISSN 00916765. URL http://www.jstor.org/stable/25148384.

B. T. Commins, R. E. Waller, and P. J. Lawther. Air pollution in diesel bus garages. *British Journal of Industrial Medicine*, 14(4):232–239, 1957. ISSN 00071072. URL http://www.jstor.org/stable/27721337.

R. Coulomb and Y. Zylberberg. Rare events and risk perception: evidence from Fukushima accident. GRI Working Papers 229, Grantham Research Institute on Climate Change and the Environment, Mar. 2016. URL https://ideas.repec.org/p/lsg/lsgwps/wp229.html.

C. C. Coutant, C. S. Wasserman, M. S. Chung, D. B. Rubin, and M. Manning. Chemistry and biological hazard of a coal ash seepage stream. *Journal (Water Pollution Control Federation)*, 50(4):747–753, 1978. ISSN 00431303. URL http://www.jstor.org/stable/25039619.

G. A. Cowman and P. C. Singer. Effect of bromide ion on haloacetic acid speciation resulting from chlorination and chloramination of aquatic humic substances. 30, 1996. doi: 10.1021/es9406905.

J. Currie and M. Neidell. Air pollution and infant health: What can we learn from california's recent experience?*. *The Quarterly Journal of Economics*, 120(3):1003–1030, 2005. doi: 10.1093/qje/120.3.1003. URL http://dx.doi.org/10.1093/qje/120.3.1003.

J. Currie and R. Walker. Traffic congestion and infant health: Evidence from e-zpass. *American Economic Journal: Applied Economics*, 3(1):65–90, January 2011. doi: 10.1257/app.3.1.65. URL http://www.aeaweb.org/articles?id=10.1257/app.3.1.65.

J. Currie, M. J. Neidell, and J. Schmieder. Air pollution and infant health: Lessons from new jersey. Working Paper 14196, National Bureau of Economic Research, July 2008. URL http://www.nber.org/papers/w14196.

J. Currie, E. A. Hanushek, E. M. Kahn, M. Neidell, and S. G. Rivkin. Does Pollution Increase School Absences? *The Review of Economics and Statistics*, 91(4):682–694, November 2009. URL https://ideas.repec.org/a/tpr/restat/v91y2009i4p682-694.html.

J. Currie, J. G. Zivin, K. Meckel, M. Neidell, and W. Schlenker. Something in the water: contaminated drinking water and infant health. *The Canadian Journal of Economics / Revue canadienne d'Economique*, 46(3):791–810, 2013. ISSN 00084085, 15405982. URL http://www.jstor.org/stable/42705901.

J. Currie, M. Greenstone, and K. Meckel. Hydraulic fracturing and infant health: New evidence from pennsylvania. *Science Advances*, 3(12), 2017. doi: 10.1126/sciadv.1603021. URL https://advances.sciencemag.org/content/3/12/e1603021.

D. Cutler and G. Miller. The role of public health improvements in health advances: The twentieth-century united states. *Demography*, 42(1):1–22, Feb 2005. ISSN 1533-7790. doi: 10.1353/dem.2005.0002. URL https://doi.org/10.1353/dem.2005.0002.

A. Davison, G. Howard, M. Stevens, P. Callan, L. Fewtrell, D. Deere, and J. Bartram. *Water Safety Plans: Managing Drinking-Water Quality from Catchment to Consumer*. 01 2005.

L. A. De Cicco, D. Lorenz, R. M. Hirsch, and W. Watkins. *dataRetrieval: R packages for discovering and retrieving water data available from U.S. federal hydrologic web services*. Reston, VA, 2018. URL https://code.usgs.gov/water/dataRetrieval.

S. Deller, A. Canto, and L. Brown. Food access, local foods, and community health. *Community Development*, 48(5):657–680, 2017. doi: 10.1080/15575330.2017.1358197. URL https://doi.org/10.1080/15575330.2017.1358197.

DHHS. Draft report on carcinogens monograph on haloacetic acids found as water disinfection by-products. Peer-review draft, US Department of Health and Human Services.

A. Ebenstein, V. Lavy, and S. Roth. The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution. *American Economic Journal: Applied Economics*, 8(4):36–65, 2016. URL https://www.aeaweb.org/articles?id=10.1257/app.20150213.

J. U. Edwards, L. Mauch, and M. R. Winkelman. Relationship of nutrition and physical activity behaviors and fitness measures to academic performance for sixth graders in a midwest city school district. *Journal of School Health*, 81(2):65–73, 2011.

EPA. Stage 1 disinfectants and disinfection byproducts rule. Technical report, EPA, 2001.

EPA. Steam electric power generating point source category: Final detailed study report. Technical report, Environmental Protection Agency, 2009.

EPA. Environmental assessment for the effluent limitations guidelines and standards for the steam electric power generating point source category. Technical Report EPA-821-R-15-006, EPA, Office of Water, Washington DC, 20460, 9 2015a. URL https://www.epa.gov/sites/production/files/2015-10/documents/steam-electric-envir_10-20-15.pdf.

EPA. Benefit and cost analysis for the effluent limitations guidelines and standards for the steam electric power generating point source category. Technical Report EPA-821-R-15-005, EPA, Office of Water, Washington DC, 20460, 9 2015b. URL https://www.epa.gov/sites/production/files/2015-10/documents/steam-electric_benefit-cost-analysis_09-29-2015.pdf.

EPA. Technical development document for the effluent limitations guidelines and standards for the steam electric power generating point source category. Technical Report EPA-821-R-15-007, EPA, Office of Water, Washington DC, 20460, 9 2015c. URL https://www.epa.gov/sites/production/files/2015-10/documents/steam-electric-tdd_10-21-15.pdf.

M. M. Fahlman, H. L. Hall, and R. Lock. Ethnic and socioeconomic comparisons of fitness, activity levels, and barriers to exercise in high school females. *Journal of School Health*, 76(1): 12–17, 2006. doi: 10.1111/j.1746-1561.2006.00061.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1746-1561.2006.00061.x.

M. Fowlie, E. Rubin, and R. Walker. Bringing Satellite-Based Air Quality Estimates Down to Earth. *AEA Papers and Proceedings*, 109:283–288, May 2019. URL https://ideas.repec.org/a/aea/apandp/v109y2019p283-88.html.

C. Freire, R. Ramos, R. Puertas, M.-J. Lopez-Espinosa, J. Julvez, I. Aguilera, F. Cruz, M.-F. Fernandez, J. Sunyer, and N. Olea. Association of traffic-related air pollution with cognitive development in children. *Journal of Epidemiology & Community Health*, 64(3):223–228, 2010. ISSN 0143-005X. doi: 10.1136/jech.2008.084574. URL https://jech.bmj.com/content/64/3/223.

S. Fretwell. Coal ash cleanup results in cleaner groundwater, greens say, 2016. URL http://www.thestate.com/news/local/article81234317.html.

A. Gambelunghe, G. Sallsten, Y. Borné, N. Forsgard, B. Hedblad, P. Nilsson, B. Fagerberg, G. Engström, and L. Barregard. Low-level exposure to lead, blood pressure, and hypertension in a population-based cohort. *Environmental Research*, 149:157 – 163, 2016. ISSN 0013-9351. doi: https://doi.org/10.1016/j.envres.2016.05.015. URL http://www.sciencedirect.com/science/article/pii/S0013935116301876.

W. J. Gauderman, E. Avol, F. Lurmann, N. Kuenzli, F. Gilliland, J. Peters, and R. McConnell. Childhood asthma and exposure to traffic and nitrogen dioxide. *Epidemiology*, 16(6):737–743, 2005. ISSN 10443983. URL http://www.jstor.org/stable/20486137.

L. Gazze. The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates. Nber working papers, National Bureau of Economic Research, Inc, Dec. 2015. URL file:///C:/Users/AYSPS/Downloads/ThePriceOfASafeHomeLeadAbatement_preview%20(2).pdf.

N. Gendron-Carrier, M. Gonzalez-Navarro, S. Polloni, and M. A. Turner. Subways and Urban Air Pollution. NBER Working Papers 24183, National Bureau of Economic Research, Inc, Jan. 2018. URL https://ideas.repec.org/p/nbr/nberwo/24183.html.

R. B. Gillespie and P. C. Baumann. Effects of high tissue concentrations of selenium on reproduction by bluegills. *Transactions of the American Fisheries Society*, 115(2):208–213, 1986. doi: 10.1577/1548-8659(1986)115⟨208:EOHTCO⟩2.0.CO;2. URL https://afspubs.onlinelibrary.wiley.com/doi/abs/10.1577/1548-8659%281986%29115%3C208%3AEOHTCO%3E2.0.CO%3B2.

A. R. Gollakota, V. Volli, and C.-M. Shu. Progressive utilisation prospects of coal fly ash: A review. *Science of The Total Environment*, 672:951 – 989, 2019. ISSN 0048-9697. doi: https://doi.org/10.1016/j.scitotenv.2019.03.337. URL http://www.sciencedirect.com/science/article/pii/S0048969719313361.

J. Gordanier, O. Ozturk, B. Williams, and C. Zhan. Free lunch for all! the effect of the community eligibility provision on academic outcomes. *SSRN Electronic Journal*, 01 2019. doi: 10.2139/ssrn.3333530.

W. B. Gray and J. P. Shimshack. The Effectiveness of Environmental Monitoring and Enforcement: A Review of the Empirical Evidence. *Review of Environmental Economics and Policy*, 5(1):3–24, 05 2011. ISSN 1750-6816. doi: 10.1093/reep/req017. URL https://doi.org/10.1093/reep/req017.

M. Guxens and J. Sunyer. A review of epidemiological studies on neuropsychological effects of air pollution. *The European Journal of Medical Sciences*, 141(3):1–7, 2012. URL https://www.ncbi.nlm.nih.gov/pubmed/22252905.

J. C. Ham, J. S. Zweig, and E. Avol. Pollution, test scores and the distribution of academic achievement: Evidence from california schools 2002-2008. *Manuscript, University of Maryland*, 2014. URL http://conference.iza.org/conference_files/TAM2012/ham_j1496.pdf.

D. Hammond, M. M. Lalor, and S. Jones. In-vehicle measurement of particle number concentrations on school buses equipped with diesel retrofits. *Water, Air, and Soil Pollution*, 179:217–225, 02 2007.

P. Handke. Trihalomethane speciation and the relationship to elevated total dissolved solid concentrations affecting drinking water quality at systems utilizing the monongahela river as a primary source during the 3rd and 4th quarters of 2008. Technical report, 2009. URL http://fliphtml5.com/jpml/bzki/basic.

B. Harder. School buses spew pollution into young lungs. *Science News*, 167(21):334–334, 2005. ISSN 00368423. URL http://www.jstor.org/stable/4016279.

G. He and J. Perloff. Surface Water Quality and Infant Mortality in China. HKUST IEMS Working Paper Series 2016-32, HKUST Institute for Emerging Market Studies, June 2016. URL https://ideas.repec.org/p/hku/wpaper/201632.html.

G. H. Heinz and D. J. Hoffman. Methylmercury chloride and selenomethionine interactions on health and reproduction in mallards. *Environmental Toxicology and Chemistry*, 17(2):139–145, 1998. doi: 10.1002/etc.5620170202. URL https://setac.onlinelibrary.wiley.com/doi/abs/10.1002/etc.5620170202.

L. Heller-Grossman, J. Manka, B. Limoni-Relis, and M. Rebhun. Formation and distribution of haloacetic acids, thm and tox in chlorination of bromide-rich lake water. *Water Research*, 27 (8):1323 – 1331, 1993. ISSN 0043-1354. doi: https://doi.org/10.1016/0043-1354(93)90219-8. URL http://www.sciencedirect.com/science/article/pii/0043135493902198.

J. A. Hoffman, E. M. Schmidt, C. Wirth, S. Johnson, S. A. Sobell, K. Pelissier, D. M. Harris, and B. T. Izumi. Farm to preschool: The state of the research literature and a snapshot of national practice. *Journal of Hunger & Environmental Nutrition*, 12(4):443–465, 2017. doi: 10.1080/19320248.2016.1227747. URL https://doi.org/10.1080/19320248.2016.1227747.

J. H. Holland, O. M. Thompson, H. H. Godwin, N. M. Pavlovich, and K. B. Stewart. Farm-to-school programming in south carolina: An economic impact projection analysis. *Journal of Hunger & Environmental Nutrition*, 10(4):526–538, 2015. doi: 10.1080/19320248.2014.980045. URL https://doi.org/10.1080/19320248.2014.980045.

W. A. Hopkins, J. H. Roe, J. W. Snodgrass, B. P. Staub, B. P. Jackson, and J. D. Congdon. Effects of chronic dietary exposure to trace elements on banded water snakes (nerodia fasciata). *Environmental Toxicology and Chemistry*, 21(5):906–913, 2002. doi: 10.1002/etc.5620210504. URL https://setac.onlinelibrary.wiley.com/doi/abs/10.1002/etc.5620210504.

S. Hu, J. D. Herner, M. Shafer, W. Robertson, J. J. Schauer, H. Dwyer, J. Collins, T. Huai, and A. Ayala. Metals emitted from heavy-duty diesel vehicles equipped with advanced pm and nox emission controls. *Atmospheric Environment*, 43(18):2950 – 2959, 2009. ISSN 1352-2310. doi: https://doi.org/10.1016/j.atmosenv.2009.02.052. URL http://www.sciencedirect.com/science/article/pii/S1352231009001800.

I. A. A. Ibrahim. Chemical characterization and mobility of metal species in fly ash–water system. *Water Science*, 29(2):109 – 122, 2015. ISSN 1110-4929. doi: https://doi.org/10.1016/j.wsj.2015.10.001. URL http://www.sciencedirect.com/science/article/pii/S1110492915000314.

Institute of Medicine. *Fitness Measures and Health Outcomes in Youth*. The National Academies Press, Washington, DC, 2012. ISBN 978-0-309-26284-2. doi: 10.17226/13483. URL https://www.nap.edu/catalog/13483/fitness-measures-and-health-outcomes-in-youth.

M. Izquierdo and X. Querol. Leaching behaviour of elements from coal combustion fly ash: An overview. *International Journal of Coal Geology*, 94:54 – 66, 2012. ISSN 0166-5162. doi: https://doi.org/10.1016/j.coal.2011.10.006. URL http://www.sciencedirect.com/science/article/pii/S0166516211002230. Minerals and Trace Elements in Coal.

J. Jalan and M. Ravallion. Does piped water reduce diarrhea for children in rural India? *Journal of Econometrics*, 112(1):153–173, January 2003. URL https://ideas.repec.org/a/eee/econom/v112y2003i1p153-173.html.

A. Jha and N. Z. Muller. Handle with care: The local air pollution costs of coal storage. w23417: 1–64, 2017. URL https://www.eenews.net/assets/2017/07/05/document_pm_01.pdf.

Y. Jiang, J. Yang, D. Cocker, G. Karavalakis, K. C. Johnson, and T. D. Durbin. Characterizing emission rates of regulated pollutants from model year 2012+ heavy-duty diesel vehicles equipped with dpf and scr systems. *Science of The Total Environment*, 619-620:765 – 771, 2018. ISSN 0048-9697. doi: https://doi.org/10.1016/j.scitotenv.2017.11.120. URL http://www.sciencedirect.com/science/article/pii/S0048969717331789.

A. Joshi, A. M. Azuma, and G. Feenstra. Do farm-to-school programs make a difference? findings and future research needs. *Journal of Hunger & Environmental Nutrition*, 3(2-3):229–246, 2008. doi: 10.1080/19320240802244025. URL https://doi.org/10.1080/19320240802244025.

D. A. Keiser and J. S. Shapiro. Consequences of the Clean Water Act and the Demand for Water

Quality. Working Papers 17-07, Center for Economic Studies, U.S. Census Bureau, Jan. 2017. URL https://ideas.repec.org/p/cen/wpaper/17-07.html.

D. A. Keiser and J. S. Shapiro. Consequences of the Clean Water Act and the Demand for Water Quality*. *The Quarterly Journal of Economics*, 134(1):349–396, 09 2018. ISSN 0033-5533. doi: 10.1093/qje/qjy019. URL https://doi.org/10.1093/qje/qjy019.

D. A. Keiser, C. L. Kling, and J. S. Shapiro. The low but uncertain measured benefits of us water quality policy. *Proceedings of the National Academy of Sciences*, 116(12):5262–5269, 2019. ISSN 0027-8424. doi: 10.1073/pnas.1802870115. URL https://www.pnas.org/content/116/12/5262.

S. Kodama, K. Saito, S. Tanaka, M. Maki, Y. Yachi, M. Asumi, A. Sugawara, K. Totsuka, H. Shimano, Y. Ohashi, et al. Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis. *JAMA*, 301(19):2024–2035, 2009.

D. A. Kopsick and E. E. Angino. Effect of leachate solutions from fly and bottom ash on groundwater quality. *Journal of Hydrology*, 54(1):341 – 356, 1981. ISSN 0022-1694. doi: https://doi.org/10.1016/0022-1694(81)90167-0. URL http://www.sciencedirect.com/science/article/pii/0022169481901670. Symposium on Geochemistry of Groundwater.

J. Kravchenko and K. Lyerly. The impact of coal-powered electrical plants and coal ash impoundments on the health of residential communities. *North Carolina Medical Journal*, 79:289 – 300, 2018. ISSN 0095-0696. doi: 10.18043/ncm.79.5.289. URL http://www.ncmedicaljournal.com/content/79/5/289.full#cited-by.

C. L. Carlson and D. C. Adriano. Environmental impacts of coal combustion residues. *Journal of Environmental Quality - J ENVIRON QUAL*, 22:227–242, 01 2009. doi: 10.2134/jeq1993.00472425002200020002x.

S. G. Lakoski, B. L. Willis, C. E. Barlow, D. Leonard, A. Gao, N. B. Radford, S. W. Farrell, P. S. Douglas, J. D. Berry, L. F. DeFina, et al. Midlife cardiorespiratory fitness, incident cancer, and survival after cancer in men: the cooper center longitudinal study. *JAMA oncology*, 1(2): 231–237, 2015.

V. Lavy, A. Ebenstein, and S. Roth. The impact of short term exposure to ambient air pollution on cognitive performance and human capital formation. Working Paper 20648, National Bureau of Economic Research, October 2014. URL http://www.nber.org/papers/w20648.

M. L. Lee, D. W. Later, D. K. Rollins, D. J. Eatough, and L. D. Hansen. Dimethyl and monomethyl sulfate: Presence in coal fly ash and airborne particulate matter. *Science*, 207(4427):186–188, 1980. ISSN 00368075, 10959203. URL http://www.jstor.org/stable/1683906.

F. Li, E. S. Lee, J. Liu, and Y. Zhu. Predicting self-pollution inside school buses using a cfd and multi-zone coupled model. *Atmospheric Environment*, 107:16–23, 2015.

L. Liang and P. C Singer. Factors influencing the formation and relative distribution of haloacetic acids and trihalomethanes in drinking water. *Environmental science technology*, 37:2920–8, 08 2003. doi: 10.1021/es026230q.

J. Liu, B. Zheng, H. V. Aposhian, Y. Zhou, M.-L. Chen, A. Zhang, and M. P. Waalkes. Chronic arsenic poisoning from burning high-arsenic-containing coal in guizhou, china. *Environmental Health Perspectives*, 110(2):119–122, 2002. ISSN 00916765. URL http://www.jstor.org/stable/3455366.

S. MacBride. The archeology of coal ash: An industrial-urban solid waste at the dawn of the hydrocarbon economy. *IA. The Journal of the Society for Industrial Archeology*, 39(1/2):23–39, 2013. ISSN 01601040. URL http://www.jstor.org/stable/43958425.

D. E. Marcotte. Something in the air? air quality and children's educational outcomes. *Economics of Education Review*, 56:141 – 151, 2017. ISSN 0272-7757. doi: https://doi.

org/10.1016/j.econedurev.2016.12.003. URL http://www.sciencedirect.com/science/article/pii/S0272775716303703.

M. Marcus. Testing the water: Drinking water quality, public notification, and school outcomes. Working paper, 2019.

J. D. Marshall and E. Behrentz. Vehicle self-pollution intake fraction: children's exposure to school bus emissions. *Environmental Science & Technology*, 39(8):2559–2563, 2005. doi: 10.1021/es040377v. URL https://doi.org/10.1021/es040377v. PMID: 15884349.

M. L. Miranda, D. Kim, M. A. O. Galeano, C. J. Paul, A. P. Hull, and S. P. Morgan. The relationship between early childhood blood lead levels and performance on end-of-grade tests. *Environ Health Perspect*, page 1242–1247, August 2007.

P. Monahan. School Bus Pollution Report Card 2006 : Grading the States. Report, Union of Concerned Scientists, May 2006. URL http://www.ucsusa.org/sites/default/files/legacy/assets/documents/clean_vehicles/pollution-report-card-2006.pdf.

N. Z. Muller, R. Mendelsohn, and W. Nordhaus. Environmental accounting for pollution in the united states economy. *The American Economic Review*, 101(5):1649–1675, 2011. ISSN 00028282. URL http://www.jstor.org/stable/23045618.

M. E. Munawer. Human health and environmental impacts of coal combustion and post-combustion wastes. *Journal of Sustainable Mining*, 17(2):87 – 96, 2018. ISSN 2300-3960. doi: https://doi.org/10.1016/j.jsm.2017.12.007. URL http://www.sciencedirect.com/science/article/pii/S2300396017300551.

T. Murray, J. Eldridge, P. Silvius, E. Silvius, and W. G. Squires. Fitnessgram® friday: A middle school physical activity and fitness intervention. internationa. *Journal of Exercise Science*, 5(1): 4–15, 2012. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4738980/.

V. Muzyka, S. Veimer, and N. Shmidt. Particle-bound benzene from diesel engine exhaust. *Scandinavian Journal of Work, Environment Health*, 24(6):481–485, 1998. ISSN 03553140, 1795990X. URL http://www.jstor.org/stable/40966811.

C. Neffand, M. Schock, and J. Marden. Relationships between water quality and corrosion of plumbing materials in buildings volume i. galvanized steel and copper plumbing systems. Sws contract report 416-i, Illinois State Water Survey Division, March 1987. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.620.7114&rep=rep1&type=pdf.

N. S. Ngo. Analyzing the relationship between bus pollution policies and morbidity using a quasi-experiment. *Medicine*, 94(37), 2015.

N. S. Ngo. Emission standards, public transit, and infant health. *Journal of policy analysis and management*, 36 4:773–89, 2017.

J. K. O'Hara and M. C. Benson. The impact of local agricultural production on farm to school expenditures. *Renewable Agriculture and Food Systems*, 34(3):216–225, 2019. doi: 10.1017/S1742170517000552.

F. Ortega, J. Ruiz, M. Castillo, and M. Sjöström. Physical fitness in childhood and adolescence: a powerful marker of health. *International journal of obesity*, 32(1):1, 2008.

C. Osmond and D. J. P. Barker. Ischaemic heart disease in england and wales around the year 2000. *Journal of Epidemiology and Community Health (1979-)*, 45(1):71–72, 1991. ISSN 0143005X, 14702738. URL http://www.jstor.org/stable/25567133.

G. Pershagen, N. Hammar, and E. Vartiainen. Respiratory symptoms and annoyance in the vicinity of coal-fired plants. *Environmental Health Perspectives*, 70:239–245, 1986. ISSN 00916765. URL http://www.jstor.org/stable/3430360.

C. Persico, D. Figlio, and J. Roth. Inequality before birth: The developmental consequences of environmental toxicants. (22263), May 2016. doi: 10.3386/w22263. URL http://www.nber.org/papers/w22263.

E. Peterson and J. V. Hoef. Stars: An arcgis toolset used to calculate the spatial information needed to fit spatial statistical models to stream network data. *Journal of Statistical Software, Articles*, 56(2):1–17, 2014. ISSN 1548-7660. doi: 10.18637/jss.v056.i02. URL https://www.jstatsoft.org/v056/i02.

E. Peterson, D. M. Theobald, and J. M. Ver Hoef. Geostatistical modelling on stream networks: developing valid covariance matrices based on hydrologic distance and stream flow. *Freshwater Biology*, 52(2):267–279, 2007. doi: 10.1111/j.1365-2427.2006.01686.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2427.2006.01686.x.

S. Petrou. Economic consequences of preterm birth and low birthweight. *BJOG: International Journal of Obstetrics and Gynaecology*, 110(20):17–23, 2003. doi: 10.18637/jss.v056.i02. URL https://www.ncbi.nlm.nih.gov/pubmed/12763106.

K. Pieper, L.-A. Krometis, B. L. Benham, and D. L. Gallagher. Simultaneous influence of geology and system design on drinking water quality in private systems. *National Environmental Health Association*, 79(2):E1–E9, 2016. URL https://neha.org/node/58494.

J. Raffo. MATCHIT: Stata module to match two datasets based on similar text patterns. Statistical Software Components, Boston College Department of Economics, Apr. 2015. URL https://ideas.repec.org/c/boc/bocode/s457992.html.

E. K. Read, L. Carr, L. De Cicco, H. A. Dugan, P. C. Hanson, J. A. Hart, J. Kreft, J. S. Read, and L. A. Winslow. Water quality data for national-scale aquatic research: The water quality portal. *Water Resources Research*, 53(2):1735–1745, 2017. doi: 10.1002/2016WR019993. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR019993.

R. Reichardt. *The Cost of Class Size Reduction: Advice for Policy Makers*. Ph.D. dissertation, RAND Corporation, 2000. URL https://www.rand.org/pubs/rgs_dissertations/RGSD156.html.

J. K. Rice. The impact of teacher experience examining the evidence and policy implications. *CALDER*, Brief 11, 2010. URL https://www.urban.org/sites/default/files/publication/33321/1001455-The-Impact-of-Teacher-Experience.PDF.

R. Ross, S. N. Blair, R. Arena, T. S. Church, J.-P. Després, B. A. Franklin, W. L. Haskell, L. A. Kaminsky, B. D. Levine, C. J. Lavie, J. Myers, J. Niebauer, R. Sallis, S. S. Sawada, X. Sui, and U. Wisløff. Importance of assessing cardiorespiratory fitness in clinical practice: A case for fitness as a clinical vital sign. *AHA*, 2016. URL https://www.ahajournals.org/doi/abs/10.1161/CIR.0000000000000461.

R. B. Russell, N. S. Green, C. A. Steiner, S. Meikle, J. L. Howse, K. Poschman, T. Dias, L. Potetz, M. J. Davidoff, K. Damus, and J. R. Petrini. Cost of hospitalization for preterm and low birth weight infants in the united states. 120(1):e1–e9, 2007. doi: 10.1542/peds.2006-2386.

A. Sappok, M. Santiago, T. Vianna, and V. Wong. Characteristics and effects of ash accumulation on diesel particulate filter performance: Rapidly aged and field aged results. *SAE Technical Paper*, 04 2009. doi: 10.4271/2009-01-1086.

C. M. Shy. Toxic substances from coal energy: An overview. *Environmental Health Perspectives*, 32:291–295, 1979. ISSN 00916765. URL http://www.jstor.org/stable/3429030.

J. Singley, B. Beaudet, and P. Markey. Corrosion manual for internal corrosion of water distribution systems. Office of Drinking Water IWH-550) EPA 570/9-84-001, U.S. ENVIRONMENTAL PROTECTION AGENCY, 1984. URL https://www.osti.gov/biblio/6808456-corrosion-manual-internal-corrosion-water-distribution-systems.

H. Soll-Johanning, E. Bach, J. H. Olsen, and F. Tüchsen. Cancer incidence in urban bus drivers and tramway employees: A retrospective cohort study. *Occupational and Environmental Medicine*, 55(9):594–598, 1998. ISSN 13510711, 14707926. URL http://www.jstor.org/stable/27730988.

E. Stets, C. Lee, D. Lytle, and M. Schock. Increasing chloride in rivers of the conterminous u.s. and linkages to potential corrosivity and lead action level exceedances in drinking water. 613-614: 1498–1509, 2012. URL https://pubs.er.usgs.gov/publication/70200036.

T. Stevens, W. Cheng, I. Jaspers, and M. Madden. Effect of short-term exposure to diesel exhaust particles and carboxylic acids on mitochondrial membrane disruption in airway epithelial cells. 181:A1031–A1031, 05 2010.

J. Sunyer, M. Esnaola, M. Alvarez-Pedrerol, J. Forns, I. Rivas, M. López-Vicente, E. Suades-González, M. Foraster, R. Garcia-Esteban, X. Basagaña, M. Viana, M. Cirach, T. Moreno, A. Alastuey, N. Sebastian-Galles, M. Nieuwenhuijsen, and X. Querol. Association between traffic-related air pollution in schools and cognitive development in primary school children: A prospective cohort study. *PLOS Medicine*, 12(3):1–24, 03 2015. doi: 10.1371/journal.pmed.1001792. URL https://doi.org/10.1371/journal.pmed.1001792.

J. Tate, R. Mason, and L. Schmitt. 2050 - the air quality emissions and health benefits of cleaner buses: A city of york (uk) case study using micro-scale models and a health impact toolkit. *Journal of Transport Health*, 5:S41, 2017. ISSN 2214-1405. doi: https://doi.org/10.1016/j.jth.2017.05.329. URL http://www.sciencedirect.com/science/article/pii/S2214140517304887.

O. M. Thompson, M. P. Twomey, M. A. Hemphill, K. Keene, N. Seibert, D. J. Harrison, and K. B. Stewart. Farm to school program participation: An emerging market for small or limited-resource farmers? *Journal of Hunger & Environmental Nutrition*, 9(1):33–47, 2014. doi: 10.1080/19320248.2013.873008. URL https://doi.org/10.1080/19320248.2013.873008.

W. Troesken. Lead water pipes and infant mortality at the turn of the twentieth century. *The Journal of Human Resources*, 43(3):553–575, 2008. ISSN 0022166X. URL http://www.jstor.org/stable/40057359.

USDA. 2015 farm to school census, 2017. URL https://farmtoschoolcensus.fns.usda.gov/data-explorer.

A. van Donkelaar, R. Martin, C. Li, and R. T Burnett. Regional estimates of chemical composition of fine particulate matter using a combined geoscience-statistical method with information from satellites, models, and monitors. *Environmental Science Technology*, 53, 01 2019. doi: 10.1021/acs.est.8b06392.

C. Villanueva, K. Cantor, S. Cordier, J. Jaakkola, W. King, C. Lynch, S. Porru, and M. Kogevinas. Disinfection byproducts and bladder cancer: A pooled analysis. *Epidemiology*, 15(3):357–367, 2004. ISSN 10443983. URL http://www.jstor.org/stable/20485906.

R. E. Waller, L. Hampton, and P. J. Lawther. A further study of air pollution in diesel bus garages. *British Journal of Industrial Medicine*, 42(12):824–830, 1985. ISSN 00071072. URL http://www.jstor.org/stable/27726103.

H. Wang, S. Masters, Y. Hong, J. Stallings, J. Falkinham, M. A Edwards, and A. Pruden. Effect of disinfectant, water age, and pipe material on occurrence and persistence of legionella, mycobacteria, pseudomonas aeruginosa, and two amoebas. *Environmental science technology*, 46, 10 2012. doi: 10.1021/es303212a.

J. Watson, D. Treadwell, and R. Bucklin. Economic analysis of local food procurement in southwest florida's farm to school programs. *Journal of Agriculture, Food Systems, and Community Development*, 8(3):61–84, Nov. 2018. doi: 10.5304/jafscd.2018.083.011. URL https://foodsystemsjournal.org/index.php/fsj/article/view/644.

G. J. Welk, A. W. Jackson, J. R. M. Jr., W. H. Haskell, M. D. Meredith, and K. H. Cooper. The association of health-related fitness with indicators of academic performance in texas schools. *Research Quarterly for Exercise and Sport*, 81(sup3):S16–S23, 2010. doi: 10.1080/02701367.2010.10599690. URL https://doi.org/10.1080/02701367.2010.10599690. PMID: 21049834.

W. Xu, G. Mai, Q. Zhu, Z. Yu, and Y. Liu. Pollution exposure at bus commuter stations in guangzhou, china. *International Journal of Environmental Technology and Management*, 19(2):103–119, 2016. URL http://EconPapers.repec.org/RePEc:ids:ijetma:v:19:y:2016:i:2:p:103-119.

Z. Yao, X. Ji, P. Sarker, J. Tang, L. Ge, M. Xia, and Y. Xi. A comprehensive review on the applications of coal fly ash. *Earth-Science Reviews*, 141:105 – 121, 2015. ISSN 0012-8252. doi: https://doi.org/10.1016/j.earscirev.2014.11.016. URL http://www.sciencedirect.com/science/article/pii/S0012825214002219.

G. Yu, D. Sun, and Y. Zheng. Health effects of exposure to natural arsenic in groundwater and coal in china: An overview of occurrence. *Environmental Health Perspectives*, 115(4):636–642, 2007. ISSN 00916765. URL http://www.jstor.org/stable/4150368.

Q. Zhang and Y. Zhu. Performance of school bus retrofit systems: ultrafine particles and other vehicular pollutants. *Environmental science & technology*, 45(15):6475–6482, 2011.

W. Zheng, K. Suzuki, T. Tanaka, M. Kohama, and Z. Yamagata. Association between maternal smoking during pregnancy and low birthweight: Effects by maternal age. *PLOS ONE*, 11: e0146241, 01 2016. doi: 10.1371/journal.pone.0146241.

F. Zhu, M. Takaoka, K. Shiota, K. Oshita, and Y. Kitajima. Chloride chemical form in various types of fly ash. *Environmental science technology*, 42:3932–7, 07 2008. doi: 10.1021/es7031168.

J. G. Zivin, M. Neidell, and W. Schlenker. Water quality violations and avoidance behavior: Evidence from bottled water consumption. Working Paper 16695, National Bureau of Economic Research, January 2011. URL http://www.nber.org/papers/w16695.

M. Zuurbier, G. Hoek, M. Oldenwening, V. Lenters, K. Meliefste, P. van den Hazel, and B. Brunekreef. Commuters' exposure to particulate matter air pollution is affected by mode of transport, fuel type, and route. *Environmental Health Perspectives*, 118(6):783–789, 2010. ISSN 00916765. URL http://www.jstor.org/stable/40661594.

Appendix D

Vita

Wes Austin was born in Atlanta, Georgia, in 1990. In high school, he captained his academic team to victory in a televised trivia tournament known as High Q. After graduating in the midst of the Great Recession, Austin decided to study Economics and History at the University of Georgia to better understand the forces that drive economic inequality and recessions. He was fortunate to spend a summer abroad in Montpellier, France, acquiring a minor in French. He also elected to take many advanced mathematics and statistics courses in preparation for graduate studies in economics, all while working part-time as a dishwasher and prep cook. After graduating *cum laude* from the University of Georgia, he continued to work in kitchens in Athens, Georgia, for another year. During long kitchen shifts, he sought inspiration from audiobooks. Eventually, he returned to his hometown to pursue a PhD in Economics from Georgia State University's Andrew Young School of Policy Studies. In his application statement of purpose, Austin described his research interests in environmental justice, wealth inequality, and urban studies. During his time at the Andrew Young School, Austin interned as a research assistant in the nearby State Charter Schools Commission of Georgia. For his efforts, he was rewarded with a prestigious university fellowship and graduate assistantship with Professor Tim Sass. In this role, he provided assistance on three major education policy research projects. One project assessed the effects of a social emotional learning experiment. Another investigated the role of principals in student achievement. The last project estimated the relationship between school district value added and community upward mobility. Meanwhile, Austin was highly active in research labs in the Andrew Young School. He worked as a research assistant for the Urban Studies Institute surveying local urban farms to help create a geographic database of urban food suppliers. He also performed essential research tasks for Georgia Policy Labs, where he worked from its inception. In his own research, Austin remained committed to expanding our understanding of environmental justice. He published one study with Garth Heutel and Dan Kreisman examining how school bus diesel emissions affect student health and cognitive performance. The primary chapter of his dissertation explores how coal combustion residuals, the most common industrial water pollutant, impact local water supplies and fetal health. Lastly, understanding the intersection of research and teaching, Austin dedicated substantial energy to instructing introductory macroeconomics and masters level data science courses to the highly diverse student body of Georgia State University. In his own time, Austin enjoys reading, playing guitar, outdoor foraging, cooking enchiladas, talking about nutrition, and pretending that growing out his beard is a new look. Austin plans to continue his research in the National Center for Environmental Economics at EPA Headquarters in Washington, DC.