Hell with the Lid Off: The Origins of Environmental Inequities in America's Most Polluted City [*]

H. Spencer Banzhaf, William Mathews, Randall Walsh

July 2025

Abstract

This study examines the changing relationship between racial segregation and environmental equity in Pittsburgh from 1910 to 1940. Utilizing newly digitized historical data on the spatial distribution of air pollution in what was likely America's most polluted city, we analyze how racial disparities in exposure to air pollution evolved during this period of heightening segregation. Our findings reveal that black residents experienced significantly higher levels of pollution compared to their white counterparts and that this disparity increased over time. We identify within-city moves as a critical factor exacerbating this inequity, with black movers facing increased pollution exposure. In contrast, European immigrants, who were also initially exposed to relatively high levels of pollution, experienced declining exposure as they assimilated over this time period. We also provide evidence of the capitalization of air pollution into housing markets. Taken as a whole, our results underscore the importance of considering environmental factors in discussions of racial and economic inequalities.

JEL Codes: J1, N5, N9, Q5

[*]Banzhaf: Department of Agricultural and Resource Economics, North Carolina State University, PERC, and NBER, spencer_banzhaf@ncsu.edu; Mathews: Department of Economics, University of Pittsburgh, WIM31@pitt.edu; Walsh: Department of Economics, University of Pittsburgh and NBER, walshr@pitt.edu.
Banzhaf and Walsh gratefully acknowledge support for this research from the Washington Center for Equitable Growth. We thank Brianna Smith for research assistance. We also thank the University of Pittsburgh Archives for granting access to the smoke studies used in this analysis.

> The entire space lying between the hills was filled with blackest smoke, from out of which the hidden chimneys sent forth tongues of flame, while from the depths of the abyss came up the noise of hundreds of steam-hammers. ... It is an unprofitable business, view-hunting; but if any one would enjoy a spectacle as striking as Niagara, he may do so by simply walking up a long hill to Cliff Street in Pittsburg, and looking over into hell with the lid taken off.
>
> —James Parton, Atlantic Monthly (June, 1868)

1 Introduction

It is a well established fact that minorities in America experience pollution burdens at higher levels than their white counterparts (Banzhaf et al. 2019; Currie et al. 2023; Colmer et al. 2024). Public recognition of these demographic patterns began some forty years ago, but when and how did they first emerge? To the best of our knowledge, this question has not previously been explored, probably because of the empirical challenge of measuring pollution before the age of modern monitoring. In this paper, we overcome this challenge with new archival data on the spatial distribution of air pollution in Pittsburgh, PA between 1910 and 1940.

We find that, at the beginning of this period, white immigrants from central Europe as well as black Americans bore disproportionate pollution burdens. Over time, these white immigrants were assimilating into native-born white culture and moving into less polluted areas. Meanwhile, as black Americans moved northward during the first wave of the Great Migration, they systematically moved into more polluted neighborhoods. The Great Migration provided many Southern black migrants with new economic opportunities in Northern industrial centers. Nevertheless, after their arrival, they faced yet a new set of challenges, including growing levels of racial segregation with attending inequities in housing, education, and employment (Massey and Denton 1993; Derenoncourt et al. 2022; Akbar et al. 2023; Fishback et al. 2023; Bondy and Sager 2020; Collins and Margo 2006, 2011). Segregation into more polluted areas was another one of those challenges.

These "polluted areas" cannot be understood with today's categories. Pittsburgh in that era was arguably one of the most polluted cities the world has ever known. As illustrated by the epigraph, popular perceptions of the city were shaded by the experience of smog and falling soot, by the filth and grime. Such unpleasantries would have only been the most apparent of the consequences. Exposure to air pollution has a range of negative outcomes including health (e.g. Currie et al. 2014; Deryugina et al. 2019; Schlenker and Walker 2016), cognitive function and mental wellbeing (e.g. Bishop et al. 2023; Kioumourtzoglou et al. 2017; Lavy et al. 2014), labor productivity (e.g. Borgschulte et al. 2022; Graff Zivin and Neidell 2013; Hanna and Oliva 2015)) and crime (e.g. Burkhardt et al. 2019; Bondy and Sager 2020). In an era after the introduction of modern-scale coke and steel production but before modern abatement technology, these effects would likely have been all the greater. Confirming this intuition, recent studies of the UK have found large health effects from historical exposure to industrial smog (Beach and Hanlon 2018; Hanlon 2024). Moreover, air pollution can have long-lasting and even inter-generational impacts, both directly and indirectly through human capital (Almond et al. 2018; Colmer and Voorheis 2020; Isen et al. 2017). Thus, historical pollution disparities could have contributed to seeding a vicious cycle of disparities in human capital, health, and exposure.

Pittsburghers were well aware of these problems, and the city's Progressive-Era reformers felt called to combat them in the name of public health and welfare—and even a thriving economy (Banzhaf and Walsh 2024). Beginning in 1910, they systematically collected falling soot and ash in glass-lined cans at set locations and weighed the contents monthly. Linking these newly discovered data to micro-level census data on individuals' locations, we study how racial and ethnic disparities in exposure to air pollution evolved between 1910 and 1940—the period of maximally increasing racial segregation in Pittsburgh. One mechanism for such disparities could be the differential sorting of poorer black people into more polluted areas, if they were unable to outbid richer whites for the more desirable areas (Banzhaf et al. 2019; Banzhaf and Walsh 2008). To test for such sorting, we also obtained US Census data on housing values to see whether they are systematically lower in more polluted areas.

Our work triangulates on three literatures connected to environmental quality: the economic history of industrial pollution, household sorting by pollution levels, and within-city environmental disparities. Within this space, our work is closest to Heblich et al. (2021). Taking a "macro approach," they reconstruct historic air quality in 142 English cities around the turn of the 19th century, using data on the location of industrial chimneys and an atmospheric dispersion model. They then document the correlation between parish-level air pollution and the share of low-skill workers. Importantly, they find evidence that low-skill workers systematically tended to live downwind of the chimneys, that those spatial patterns were not present in the pre-industrial era, and yet that they continue to persist today.

In contrast, in focusing on the United States at the time of the Great Migration, we examine a very different socio-economic setting, one where racial segregation is particularly salient. Intriguingly, these relevant racial categories were shifting over time. At the beginning of the period, first-generational Eastern European immigrants were on the low rungs of Pittsburgh's socio-economic ladder and socially segregated from native-born whites. This changed as they assimilated into white America and as black migrants arrived from the rural south.

Our paper also differs from Heblich et al. (2021) in taking a "micro approach" that focuses on a single U.S. city (Pittsburgh). This approach has two advantages. First, the focus on Pittsburgh allows us to use the unique measures of sootfall gathered between 1910 and 1939, rather than modeled data.[1] Second, we can link these early measures of actual pollution to micro-level data from the full-count decadal censuses. Thus, we are able to assess the relationship between a much richer set of demographics and pollution exposure. Moreover, for a subsample, we have a panel of individuals observed over time from the Census Tree Linking Project (Price et al. 2021). Because they track individual movements, these data enable us to test for patterns of mobility. Finally, with 1930-40 housing prices from the Census, we can test for market impacts from sorting.

We find that, as early as 1910, correlations between pollution, income, race and nativity had been established in a manner that likely reinforced societal inequities. Notably, we find race and nativity to be markedly more important than income for predicting pollution exposure, a pattern still found today (Colmer et al. 2024). Furthermore, racial inequity in exposure to air pollution increased significantly between 1910 and 1940, while inequity by nativity declined. By 1940, on average, black Pittsburghers were exposed to over a one-half standard deviation more pollution than were their white counterparts of similar income, age, and marital status. We estimate this black-white disparity in air pol-

[1] To the best of our knowledge, comparable air pollution data do not exist for any other U.S. city from this time period. The archived correspondence between the Pittsburgh researchers and activists in other cities indicates that Pittsburgh was pioneering a new endeavor, along with London. We are aware of sootfall studies in London and Glasgow between 1910 and 1912, Cincinnati between 1930 and 1940, and New York between 1936 and 1976, but we have been unable to locate detailed records of these data. St. Louis also attempted a similar study, but the data proved unusable and the project was abandoned.

lution to be roughly six times as large as that associated with a one standard deviation difference in income.

In contrast to twenty-first century reductions in disparities, which appear to be driven by patterns in pollution mitigation (Currie et al. 2023), these earlier increases in disparities appear to be based on black residents “coming to the nuisance” rather than pollution coming to them. Notably, from 1910 to 1940, we find black households moving to areas that were already polluted by 1910. Additionally, in micro data of within-city moves, we find them systematically moving to more polluted areas. In contrast, the movement of native-born white individuals were associated with decreasing pollution. Foreign-born movement was also associated with decreasing pollution, at about the same rate as native born whites or faster. Thus, by the end of the period, a stark difference in black-white exposure emerged.

Consistent with this sorting mechanism, we also find impacts of air pollution in Pittsburgh's housing market. For 1930-1940, a one standard deviation increase in air pollution was associated with a five to six percent reduction in housing price/rent. Nevertheless, housing prices cannot explain all the racial patterns that we observe, as black-white disparities remain even after conditioning on prices.

2 Context

On the eve of World War I, Pittsburgh was the United States' 8$^{\text{th}}$ largest city and ranked 9$^{\text{th}}$ in GDP. It especially excelled in the production of iron and steel and, as an intermediate process, the production of coke. Accordingly, it also led in the consumption of pollution-generating fossil fuels. It was second only to Chicago in annual consumption of bituminous coal (3.8 million tons vs. 5.3 million tons) and coke (1.4 million tons vs. 2.0 million tons) and led the country in annual consumption of coal gas at 16.7 billion cubic feet.[2] Yet Pittsburgh was a mere quarter the size of Chicago in terms of both land area and population, making its emissions much more dense.

Pittsburgh's unique geography only added fuel to the fire. It sits at the confluence of the Allegheny and Monongahela Rivers, where they form the Ohio River. To access coal and metals by barge, Pittsburgh's industries were nestled into these river valleys.

[2] Production and consumption figures come from the 1914 Census of Manufacturers (U.S. Department of Commerce 1917).

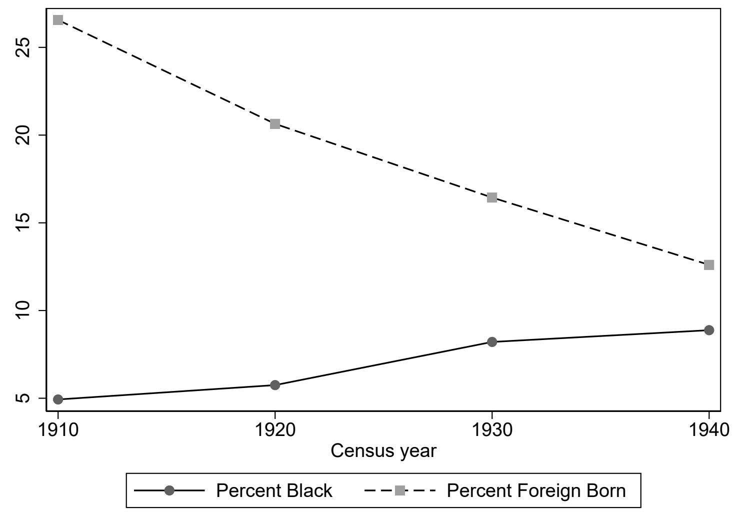
Residential areas, in contrast, were on the sides and top of steep hills rising up from the rivers. This geography has a two-fold significance for our study. First, it implies pollution patterns were exogenous to downwind demographics. That is, polluters located where they were because of where the river valleys were, not because of who lived downwind, while residents could sort in response, with richer or more privileged groups obtaining access to the cleaner high-elevation areas. Second, it implies temperature inversions often would trap coal smoke for days at a time in the valleys. Combined with its dense emissions of smoke, these conditions created what was, at this time, almost certainly the United State's most polluted city—hence, James Parton's epigraph for Pittsburgh, "Hell with the lid taken off."

Pittsburgh's smoke investigators, led by John O'Connor Jr., an economist, were trying to address this problem. They researched abatement technologies and lobbied for reforms. They also tried to document the extent of the problem. While collecting the sootfall data, they researched smoke's damages, estimating concentration-response functions for various injuries and resulting monetary losses, in what was essentially an early integrated assessment model (O'Connor 1913). They estimated damages from higher laundry bills from the soot, higher cleaning costs, damaged textiles and merchandise, and so forth. Importantly, they were convinced that the smoke problem was common knowledge to Pittsburghers. The effects were "axiomatic" or "nobody would deny" them, they would write. They also observed that people were intentionally adapting to the problem. For example, Pittsburgh was known as "the mourning town" because the men wore dark suits to avoid showing dirt. Women replaced woolens with more washable fabrics, and people used washable paints for their homes and were less likely to use wallpaper. Because they were like a "tax" that one had to pay to live in Pittsburgh, said the investigators, these nuisances kept some people away, driving up wages and putting the city at a competitive disadvantage (O'Connor 1913).[3]

Moreover, according to O'Connor, people were well aware of which neighborhoods were dirtier than others, so housing demand and, hence, prices were lower in the dirtier neighborhoods. He based this conclusion on the fact that Philadelphia tax assessors decreased their assessments in the two dirtiest wards of that city, on court decisions awarding damages to injured property, and the expert opinion of Pittsburgh's real estate firms, which would sometimes advertise property as free from smoke (O'Connor

[3]For a broader discussion of the sootfall studies and estimates of damages, see Banzhaf and Walsh (2024).

Figure 1: Demographic Trends in Pittsburgh, 1910-1940



Note: This figure shows the evolution of demographic groups in Pittsburgh. The solid black line is the share of the black population and the dashed line is the share of foreign-born.

1913).[4] People also understood that poorer and disadvantaged people lived in the dirtiest neighborhoods. In 1941, when Pittsburgh's city council finally passed binding rules to reduce its smoke problem, largely by targeting home heating technologies, they wondered whether it would buy them or cost them votes in poorer neighborhoods, as the air quality improvements would benefit the poor neighborhoods but the costs of the clean fuels would be regressive (Tarr and Lamperes 1981). In summary, the contemporary wisdom was that people were sorting in response to pollution.

Meanwhile, like virtually every other large Northern city in the United States over the 30 years that we study, Pittsburgh experienced significant demographic change. Its population grew by over 25% between 1910 and 1940, from 533 thousand to 672 thousand. As seen in Figure 1, at the beginning of this period, its population was only 5% black but 27% foreign-born. Comprising the largest disadvantaged group at this time, many of these immigrants were from Eastern Europe and many did not speak English. However, over time these population shares converged, as European migrants assimilated and more black households moved from the South. This convergence was the result of two factors. First, during the first wave of the Great Migration, Pittsburgh's black

[4] Intriguingly, earlier the prestigious economist John R. Commons had proposed a survey of quality-adjusted housing costs in Pittsburgh's different neighborhoods. He submitted smoke as one factor to consider in the quality adjustments (Commons 1908).

Figure 2: The Rise of Racial Segregation in Pittsburgh

Note: This graph shows the rapid increase in Pittsburgh's racial segregation between 1910 and 1940, using two measures of segregation. The solid black line is the Isolation Index and the dashed gray line is the Dissimilarity Index. (See Shertzer et al. (2016) for a discussion of the underlying data.)

population more than doubled, increasing from 26 thousand to 62 thousand. From 1910 to 1940, the share of black residents increased from 5% to 9%. As was typical of other other large Northern and Midwestern U.S. cities over this period, these increased rates of black migration were met with accelerating waves of white flight (see Shertzer and Walsh 2019, for a discussion). Accordingly, as with other U.S. cities, 1910 to 1940 marked the period of maximal segregation growth for Pittsburgh. Figure 2 shows two measures of segregation over time, the isolation index and the dissimilarity index. The isolation index measures the percent black in the neighborhood of the average black resident. The dissimilarity index measures the share of the black (or white) population that would need to relocate in order for both races to be evenly distributed across a city. As shown in the figure, both measures rapidly increased from 1910-40.

Second, among whites, patterns of nationality were also changing over this time period. European immigration to the U.S. slowed drastically during World War I. Then, after a brief uptick following the war, it slowed again as a result of federal restrictions on immigration flows under the quota acts of 1921 and 1924 (Abramitzky and Boustan 2017). Consequently, Pittsburgh's foreign-born population declined sharply over this time period, from 141 thousand in 1910 to 85 thousand in 1940. As a percent of the population, levels fell by half from 27 percent to 13 percent. Further, not only were their numbers greatly reduced, but this smaller group of foreign-born Pittsburghers would have, on average, been in the country for a longer period of time. This last point is important because, as we will show below, assimilation (speaking English) was associated with significantly reduced inequity in pollution exposure for foreign-born Pittsburghers.

3 Data

We utilize data from a several sources. Air pollution data come primarily from a novel set of "Sootfall" measurements that were collected over several decades in the city of Pittsburgh. The main sources of demographic and income data are the full-count U.S. Decennial Censuses, which we augment with data on the incomes of county employees taken from annual reports of the Allegheny County Controller's Office. We also take advantage of the Census Tree Linking Project (Price et al. 2021) to track individuals across census waves. Finally, for our analysis of housing markets, we match house characteristics from current day Assessor's records to self-reported house values and rents from the 1930 and 1940 Censuses.

3.1 Sootfall Data

Our primary measure of air pollution comes from the series of "smoke studies" introduced above. These studies were conducted by collecting ash-fall in glass-lined metal cans placed on rooftops throughout the city. The contents were weighed monthly with the results recorded in tons of soot per square mile per month.

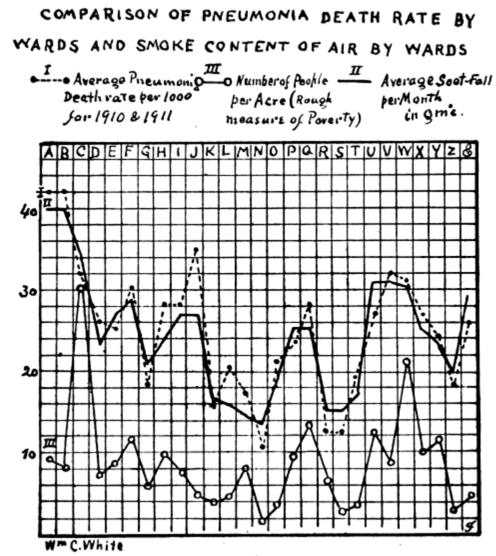
The first study, published by White and Marcy (1912), was conducted in 1910 at the University of Pittsburgh with funding from Richard K. Mellon. The researchers collected one year's worth of spatially delineated sootfall data during the year and reported annual averages for each of Pittsburgh's 26 Wards. Figure 3 reproduces two key figures from their report. Panel A records the spatial distribution of sootfall in the city (with darker shading corresponding to higher levels of sootfall). Panel B reports measured sootfall levels along with Ward-Level data on pneumonia death rates and population density. This second panel highlights one main motivation for the sootfall studies, namely to assess the link between smoke and mortality.

The remaining studies reported monthly data at discrete locations. These include

Figure 3: Early Sootfall Measurement in Pittsburgh

(a) Ward-Level Sootfall

(b) Pneumonia and Sootfall



Note: Figures reproduced from: White, Charles William and C.H. Marcy, "A Study of The influences of Varying Densities of City Smoke on The Mortality From Pneumonia and Tuberculosis" in Transactions of the Fifteenth International Congress on Hygiene and Demography (1912)

studies in 1912-13, 1923-24, and 1929-30 by the Mellon Institute of Research.[5] Finally, the City of Pittsburgh's Bureau of Smoke Regulation (part of the Department of Public Health) undertook a large scale sootfall study during the years 1938 - 1939 with support from the Works Progress Administration.

Table 1 reports the duration, scope, and organization responsible for each sootfall study in our data set. The table also displays the mean and standard deviation of pollution across sites within each study. For our analysis, we digitized each of these studies. Based on these raw data, in 1910 the average sootfall in Pittsburgh was 69.6 tons per square mile per month (or 0.29 kg/m^2/year), a figure which grew to 118.6 t/mi^2/mo (0.50 kg/m2/yr) in 1923/1924 before falling again in future years. Furthermore, there is substantial geographic variation within each time period, with coefficients of variation

[5]The Mellon institute of Research was part of the University of Pittsburgh from its inception in 1913 until 1928 when it became a not-for-profit independent research institute. It became part of Carnegie Mellon University in 1967. Data from the Mellon Institute Studies as well as the 1938 Study undertaken by Pittsburgh's Bureau of Smoke Regulation are located in the Archives of the University of Pittsburgh (Ref# US-PPiU-ais198307).

Table 1: Sootfall Studies

Study Sponsor/Author	Years	Months	No. Sites	Mean (SD)
White & Marcy (1912)	1910	Annual Avg	26 Wards	69.6 (20.9)
Mellon Institute	1912-13	12	12	89.7 (42.3)
Mellon Institute	1923-24	11	12	118.6 (51.5)
Mellon Institute	1929-30	12	19	80.8 (41.3)
Works Progress Administration	1938-39	12	100	94.6 (75.9)

Note: This table summarizes the sootfall data available, including the researchers, time period studied, and number of sites. (The 1910 study reported ward-level averages rather than sites.) The table also shows the mean and standard deviation of sootfall measured in tons / mi^2 / mo.

ranging from 0.3 to 0.8.

To put these values in perspective, a similar study in London estimated sootfall to be 36 t/mi^2/mo at the most polluted part of the city in 1910-11 while Glasgow's was 68 t/mi^2/mo (Mellon Institute 1914). Thus, Pittsburgh's pollution would appear to be much worse than Britain's industrial cities at the same time period. For additional perspective, we have calibrated these figures to modern measures of total suspended particulate pollution (TSP), using later data when modern monitors overlapped with sootfall studies. Based on that calibration, Pittsburgh's air in the first quarter of the 20th century had TSP levels on the order of 500 to 1000 μg/m^3, vs. roughly 30 today.[6] Historical estimates put London's turn-of-the-20th-century TSP levels closer to 600 μg/m^3 (Brimblecombe 1987; Fouquet 2011), while China's northern cities (the dirtiest) at the turn of the 21st century were as high as 550 μg/m^3 (Chen et al. 2013). These comparisons suggest that Pittsburgh's historical pollution was some of the worst people have ever experienced.

But what is sootfall? Simply put, it is grit in the air, or primary particulate matter from fossil fuel combustion—soot and ash—that is deposited fairly close to combustion sources. This measure of pollution will be unfamiliar to most modern readers. Nevertheless, it is highly relevant for a historical study for three reasons.

First, objectively, soot and sootfall caused significant economic damages. This claim may be surprising given our current understanding that fine particles penetrate deeper into the lungs than course particles, causing worse health effects. However, the relative

[6] In 1959, the last sootfall study overlapped with the first of the modern TSP monitors. Comparing the citywide monthly average sootfall to the monthly reading of the TSP monitor (which only operated 1-3 days per month), we get an average TSP-to-sootfall ratio of 7.2:1. Using only the nearest sootfall station instead of the annual average, we get a ratio of 8.6:1. The next best match we can find is in Manhattan, where the annual average TSP reading in 1936 was 450 μg/m^3 and the 1932-34 average sootfall was 110 t/mi^2/mo, for a ratio of 4.1:1 (Eisenbud 1978).

damages caused by fine particles today must be understood to hold in a context where we have already picked the "low hanging fruit" of controlling the coarsest particles. By contrast, in prewar Pittsburgh, there was no effective zoning constraining the location of factories, combustion was incomplete relative to today's processes, and chimneys were typically at lower heights, with no baghouses or other filters. According to chemical analyses at the time, the resulting sootfall deposits consisted largely of ash, fixed carbon, and iron oxide, but also with tar and molecules comprised of sulfates and ammonia (Mellon Institute 1914) (ammonium sulfate is a major component of $PM_{2.5}$ today). Soot particles fell into the eyes (a frequent complaint) and filled the nose, both pathways to health consequences. As depicted in Figure 3 above, White and Marcy (1912) found a strong spatial correlation between sootfall and pneumonia incidence (see also Mills 1943). Falling soot also came in contact with the skin. As early as 1795, contact with soot was identified as causing cancer in populations of chimney sweeps, a finding that was well known to the sootfall scholars and which has held up over time (Evanoff et al. 1993). To this day, the EPA warns people to avoid skin contact with ash and soot and to protect the eyes.[7]

In addition to such health effects, Mellon Institute researchers examined the effect of sootfall on vegetation in laboratory experiments, finding that contact with Pittsburgh's soot slowed growth rates (Clevenger 1913). Other researchers found it corroded metals (Benner 1913). It soiled clothing which had to be washed more frequently and also penetrated into buildings, where it soiled the walls and furnishings and damaged merchandise, which was often covered in glass for protection (O'Connor 1913).

Second, regardless of the objective facts, subjectively soot and sootfall were a salient factor that plausibly would have driven sorting and segregation. As noted above, research findings in circulation at the time indicated that contact with soot caused cancer. The deposition from sootfall was clearly perceptible, irritating eyes, and smudging clothes and materials. For all these reasons, sootfall measures continued in such cities as London and New York until the 1970s (Brimblecombe 1987; Eisenbud 1978). And as late as 1990s, estimates of air pollution damages still included such effects as eye irritation, soiling, and materials damages, even if mortality effects were beginning to dominate (Banzhaf et al. 1996; Rowe et al. 1996).

Third, historically sootfall was correlated with other measures of combustion and

[7] US EPA, "Wildfire Smoke Factsheet: Protect Yourself from Ash."

pollution, indicating our measures contain a high signal-to-noise ratio. A 1936 study in New York found a high spatial correlation between sootfall and solid fuel combustion (Eisenbud 1978). As a further check on the signal provided by these data, we compare them to time series data recorded in the U.S. Weather Bureau's meteorological reports for downtown Pittsburgh from 1905 to 1935. Observers reported daily whether they detected light or heavy smoke from their meteorological station in downtown Pittsburgh.[8] We aggregate these observations to the monthly level to compare them to the sootfall data. There are 41 months in which air pollution is measured by both the Mellon Institute soot studies and the U.S. Weather Bureau's daily reports for downtown Pittsburgh. To gauge whether the sootfall data comports with the Weather Bureau's reports, we regress monthly sootfall data on the number of heavy smoke days reported at the Pittsburgh Meteorological Station.

Table 2 reports the results. In the first column, the dependent variable is the city-wide average across all sootfall stations, while in the second column we only use data from the downtown monitoring site closest to the meteorological station. Although we find no statistically significant relationship between city-wide soot averages and reports of heavy smoke downtown, we do find a very strong statistical relationship between downtown sootfall and observed smoke downtown. Here, one extra heavy smoke day is associated with an increase in sootfall rate equal to 16.47 tons per square mile per month. Moreover, the meteorological data explain 43% of the variation in the sootfall data. Figure 4 provides a scatter plot of the data underlying the regression in column 2 of Table 2. These results are consistent with the idea that sootfall measures and observed smoke are both measuring local pollution. They are correlated over time, but only when we properly account for the spatial variation by looking at the sootfall monitors closest to the meteorological station instead of taking a city-wide average. We conclude that these historical sootfall data are meaningful.

3.2 Demographic Data

We merge these pollution data to demographic data from three sources, the full-count U.S. Censuses for the years 1910, 1920, 1930 and 1940; a panel of individuals linked

[8]These data are taken from "United States Weather Bureau Report of the Chief." These annual reports are available through the Carnegie Library of Pittsburgh's Archives. Beginning in 1935, the Pittsburgh Meteorological Station moved from downtown Pittsburgh to the airport in Moon Township, outside our study area. Accordingly, we limit our smoke days comparisons to pre-1935 months with sootfall data.

Table 2: Regressing Sootfall on Heavy Smoke Days

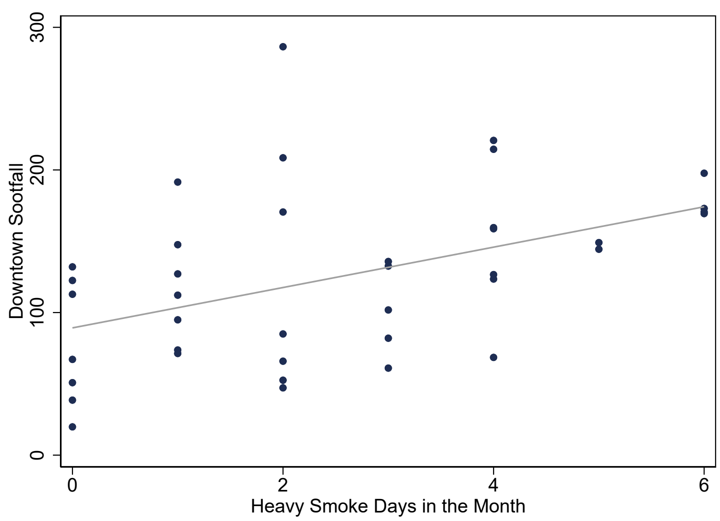
	City-Wide Soot Average	Downtown Soot
Heavy Days	-0.734	16.47***
	(4.092)	(5.092)
Constant	110.4***	123.5***
	(26.32)	(32.21)
N	40	39
R^2	0.071	0.432
adj. R^2	-0.342	0.169

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table shows results from a regression of sootfall data on the Weather Bureau's estimate of the number of smoky days in each month. The model includes month fixed effects and drops one extreme outlier (November, 1929).

Figure 4: Downtown Sootfall and Heavy Smoke Days



The figure illustrates the regression line from Table 2, Col. 2.

across Census years; and income data for public employees of Allegheny County.

3.2.1 Full-Count US Census

We obtained full-count US Census demographic data from IPUMS (Ruggles et al. 2020) for all Pittsburgh residents between the years 1910 through 1940. For each year, we use the smallest possible (digitized) census geography. Specifically, for 1910-1930 we use Enumeration District shapefiles originally digitized by Shertzer et al. (2016). For 1940, we use Census Tract shapefiles downloaded from the National Historical Geographic Information System. Enumeration Districts designate the area that a single census enumerator could survey in a two week period, typically a handful of city blocks. On average, Pittsburgh's Enumeration Districts spanned 0.11 square miles and contained 1,424 residents in 1910, 0.09 square miles and 1,319 residents in 1920, and 0.1 square miles and 1,340 residents in 1930. Constituting a higher level of aggregation, Pittsburgh's Census Tracts in the 1940 Census spanned 0.29 sq miles and held 3,010 residents on average.

These IPUMS data include micro-level information about every individual in Pittsburgh. For each individual, we obtain information on race, nativity, English speaking, age, sex, home ownership, labor force participation, occupation, marital status, and relationship to head of household. For 1930 and 1940, we also obtain tenure and self-reported rent or home value, as well as income in 1940 (these data were not available earlier). Because the Census doesn't report income prior to 1940, for 1910-30 we proxy for income using Occupational Income Scores. These scores have been in widespread use for thirty years. They are based on analysis of income, occupation and industry data from the 1950 census and backcast by Ruggles et al. (2020). Each point in the Occupational Income Score represents $100 in 1950 income. Our primary estimation sample takes only household heads (male or female) as the unit of observation. These heads can be male or female, but approximately 80-85% of household heads are male during this period. However, as discussed later, our results are not sensitive to alternative sample definitions.

Approximately 108,000 to 150,000 individual household heads are included in our main sample, depending on year. Summary statistics for this sample are presented in Appendix Table A.1. The table provides means and standard deviations for each variable, separately for each Census year. It confirms the patterns seen in Figure 1, including the rise in the black population over time and decline in foreign-born population.

3.2.2 Census Linking

We utilize the Census Linking Project (see: Price et al. 2021, 2023) to generate linked samples of individuals across the 1910 and 1940 Censuses and also across the 1930 and 1940 Censuses. Using the Census Tree linkages, we start with the universe of individuals living in Pittsburgh who appear in the 1940 Census and then link backwards to the 1910 and 1930 Censuses respectively. Because we are primarily concerned with how within-city moves contributed to patterns of pollution exposure, we further limit the samples to those individuals who were enumerated as living in Pittsburgh for both Censuses. After this filtering, we are left with 85,363 individuals linked between 1910 and 1940 and 297,565 linked individuals between 1930 and 1940.

Appendix Table A.2 provides summary statistics for linked estimation samples, as well as the baseline unlinked 1940 estimation sample for comparison. As seen in the table, linkage rates are significantly lower for black individuals. Interpreting the differences, it is important to note that some of the shrinkage in the relative size of the Black sample is mechanical, as the percentage of the Pittsburgh population that was black in 1910 and 1930 was lower than it was in 1940. The size of the Black sample in the 1930 - 1940 link is roughly 30% smaller than it would be had black and white linkage rates been identical. For the 1910 - 1940 linkages, the gap is roughly 50%. In spite of their lower linkage rates, we still have 17,081 linked Black individuals in the 1930-1940 sample and 2,191 Black individuals in the 1910-1940 sample. Mechanically, the individuals in the 1910-1940 sample are much older (12-15 years) than the other samples and hence have higher average incomes in 1940. Individuals in the linked samples are more likely to be home owners. We also link fewer women than men, but these differences are relatively small.

3.2.3 Allegheny County Employee Data

As noted previously, the Decennial Census did not begin collecting data on income until 1940, so we primarily rely on Occupational Income Scores. Although widely used in the economic history literature, Occupational Income Score is admittedly a noisy measure of income. To augment these measures and assess their validity, we also digitized data on the incomes of employees of Allegheny County (where Pittsburgh is located) taken from the 1910, 1920 and 1926 reports of the Allegheny County Controller.[9] For the 1910 and 1920 reports, we hand-matched the individuals to Census records using their names and occupations. For 1910 (1920) we were able to uniquely match 226 (437) of the 751 (1,252) individuals included in the report to individuals in the Census living inside the City of Pittsburgh. We limit attention to individuals living inside the city because all of the sootfall stations are located within the city boundary. For these two years, we recover each individual's residential location (Enumeration District), race, and nativity from their Census records. We also recover Occupational Income Score for 224 (235) of these County employees.

The 1926 report falls between Census years, so we do not attempt to match employees to Census records. Thus, although we have income from the employment records, we do not know other demographic information for these individuals (e.g., race and nativity). However, in this case we have actual addresses. Thus, we can still link them to our pollution measures. We successfully geocode the addresses for 1,381 of these 2,530 individuals.

Appendix Table A.3 presents summary statistics for matched county employees. These data can be used as an alternative estimation sample for understanding the exposure of different demographic groups to pollution, a sample that contains actual incomes rather than Occupational Income Scores.

The 1910 and 1920 data can also be used to assess the validity of using Occupational Income Scores as a proxy for income in our other samples. We gauge this validity in the following way. First, using the Bureau of Labor Statistics CPI series, we convert the 1910 and 1920 incomes to 1950 dollars. Then, separately for 1910 and 1920, we regress each individual's actual income (as reported by the County) on their Occupational Income Score (based on the occupations reported in the IPUMS version of the Decennial Census). If (i) relative wages across incomes and occupations remained unchanged between 1910/1920 and 1950; (ii) County employees are paid market wages; and (iii) the CPI adjustments to 1950 dollars are perfect, then the estimated coefficients in this regression should equal exactly 100. Here, condition (i) is essentially the hypothesis to be tested and conditions (ii) and (iii) are conditions under which the test is valid. Table 3 presents the regression results. The coefficients estimates are 107.6 and 60.12 for 1910 and 1920, respectively. For both decades the estimates have a high degree of statistical significance.

[9] 1926 was the last year for which individual salaries are reported in these reports. It is also unique in that is the only report we could find that actually listed each individual's home address.

Table 3: Regressing Income on Occupation Scores

	1910 Income	1920 Income
Occ Income Score	107.6***	60.12***
	(16.44)	(6.990)
Constant	685.3	679.1***
	(455.1)	(207.2)
N	224	235
R^2	0.162	0.241
adj. R^2	0.158	0.238

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows results from a regression of income on Occupational Income Scores. Nominal incomes are converted to 1950 Dollars using the Bureau of Labor Statistics' CPI Series. If county employees are representative and CPI adjustments are accurate, a coefficient on Occupational Income Score close to 100 indicates a good proxy.

Taken together, these estimates suggest that Occupational Income Scores are likely to be a good proxy for income in our pollution analysis.

3.3 Housing Characteristics

Beginning in 1930, the Decennial Census collected data on self-reported housing values and rents. These values can be regressed on pollution to test whether sorting processes are leading to lower housing values in more polluted areas. To control for housing characteristics in this regression, we match the household heads' addresses from the 1930 and 1940 censuses to the 2022 Allegheny County Assessor's rolls. These data have information about the housing characteristics as of 2022. In this process, we conservatively restrict the sample to single family homes that were built before 1930/1940. This insures that we do not have any teardowns and rebuilds between the 1930-40 Census data and the observed house in 2022. Again conservatively, we also restrict the sample to addresses with a single head of household. From both sides of the match, we begin by standardizing the addresses for how type of road (i.e. street, avenue, road, etc.), numbered streets (i.e. 1st avenue) and cardinal directions are delineated. We then limit the matched sample to only those matches where the street name match was completely unambiguous. Next, we drop any match whose location as identified in the Assessor's data-set lies outside its reported Enumeration District (1930) or Census Tract (1940). This step provides an evaluation of the quality of our matching procedure. For 1930, only about 4% of our matches fail this filter. For 1940, the failure rate is just under 10%. As a last step, we drop observations that have incomplete Assessor's data. Statistics for the matching process are presented in Appendix Table A.4.

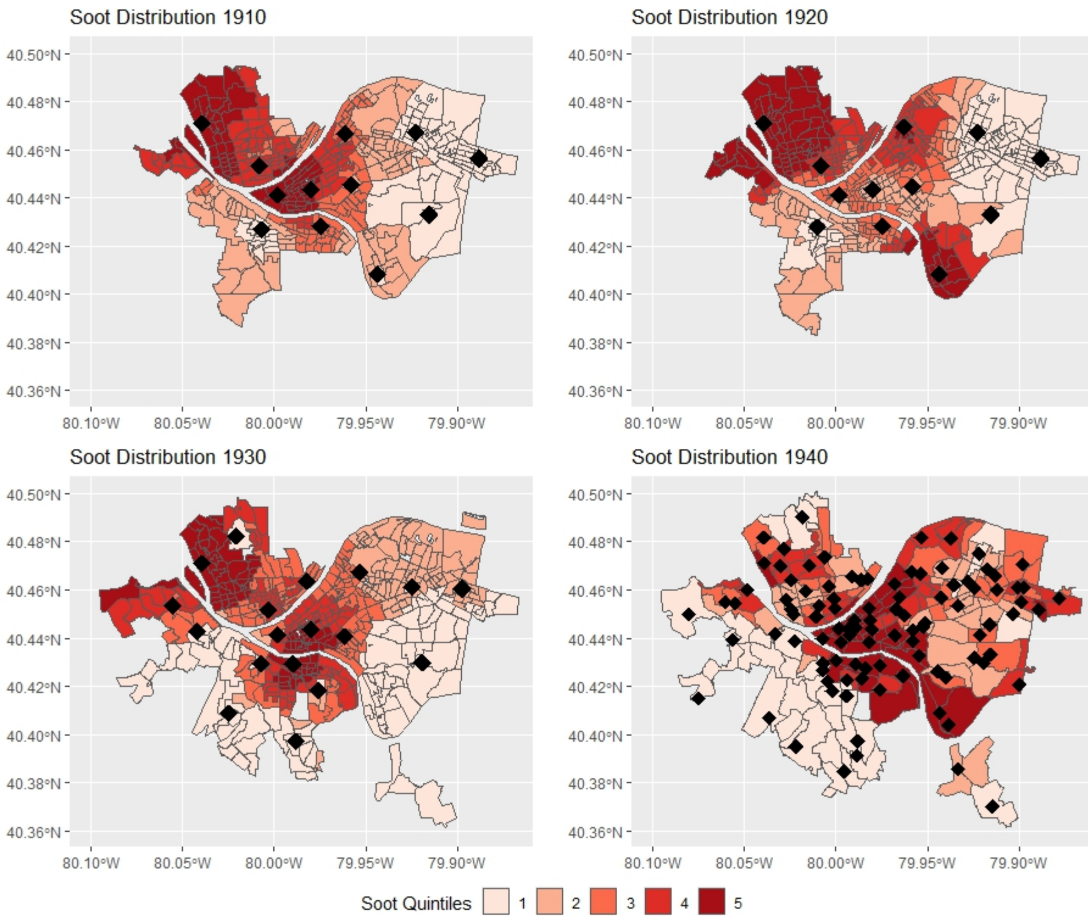
As shown in Appendix Table A.1, 40% of our sample are home owners in 1930, a share which falls during the Great Depression, to 32% in 1940. Consistent with that trend, our final sample comprises 12,635 owner-occupied homes in 1930 and 11,345 in 1940. However, despite the fact that a majority of residents are renters, because we focus on single-family houses with observable characteristics our sample of rents is smaller, at 3773 in 1930 and 6296 in 1940.

Potentially, one might be concerned about bias or measurement error in self-reported house values. However, to test this issue, Akbar et al. (2023) hand-matched a subset of 1930 and 1940 Census data to assessor records of contemporaneous transactions. They found that self-reported house values from the U.S. Census provide an unbiased estimate of actual sale prices for 1930 and 1940. Although measurement error might still be a concern, we note that, in the context of hedonic regressions, this would be measurement error in the dependent variable. Moreover, Banzhaf and Farooque (2013) have found that community-level price indices constructed from more recent self-reported valuation data are highly correlated with both price indices constructed from sales and with public goods and amenities. They also found that self-reported rents are even more accurate than self-reported values, as one might expect.

3.4 Matching Pollution to Census Data

To assign pollution data to census divisions (Enumeration Districts in 1910, 1920, and 1930 and Tracts in 1940), we follow a 3-step process. First, we collapse every station in each sootfall study to an annual average, dropping missing values. Next, limiting ourselves to pollution readings within five years of a census year, we take an inverse-distance-weighted average of the three closest stations' annual average to a census division's centroid. This step potentially leaves us with multiple pollution estimates for a census division. For example, in 1910 we have pollution estimates based on both the 1910 and 1912 soot studies. Therefore, we take an average of the different estimates for every census division. Finally, individuals are assigned to the pollution at the centroid of their division, with one exception for the 1926 county employee data where we assign pollution to the individual's known address.

Figure 5: Spatial Distribution of Sootfall as Attached to Census Geography



Heatmap of pollution after attachment to census geographies. Each diamond shows the location of individual sootfall stations. For 1920, 1930 and 1940, the interpolations are based solely on the 1923/24, 1929/1930 and 1938/39 studies respectively. For 1910, we report the average of the Ward-Level data for 1910 and the interpolated station data for 1912/13. We report the within Census quintiles of sootfall. For 1910 - 1930, sootfall is interpolated to the Enumeration District Level. For 1940 it is at the Census Tract level.

In Figure 5 we present the interpolated spatial distribution of sootfall and the location of the individual sootfall collection stations. The interpolated data are reported in terms of within-census-year quintiles. Overall, the spatial distribution of pollution is fairly stable, with pollution levels highest along and between the rivers and lower to the southwest and in the highest elevation areas to the east. However, there is one notable exception: a relative decrease in pollution that occurred in Pittsburgh's Hill District (very center of the map, between the Allegheny and Monongahela Rivers) in the 1923-24 data. At the time, the Hill District was the heart of the city's black community. This short-lived relative decrease came about as the result of a major spike in steel production during 1923 (see Appendix Figure A.2) that was associated with a prolonged period of the City's mills operating at extremely high capacities. This surge in production caused

Table 4: Correlation in Enumeration District Sootfall Across Time

	Sootfall 1910	Sootfall 1920	Sootfall 1930	Sootfall 1940
Sootfall 1910	1			
Sootfall 1920	0.448	1		
Sootfall 1930	0.745	0.467	1	
Sootfall 1940	0.268	0.0301	0.258	1

Note: Correlation across years and Enumeration District for sootfall as attached to the 1920 geography.

Table 5: Average Sootfall Exposure by Year and Demographic Group

	Demographic Group		
	Black	Foreign Born	White
Year			
1910	85.0	86.3	80.0
1920	108.8	116.3	116.0
1930	91.5	86.5	82.9
1940	116.3	100.0	85.9

This table shows average sootfall in each year, after its assignment to Census geographies, as experienced by different groups. Note: White includes all non-foreign-born white individuals, foreign-born includes all non-black foreign-born individuals and black includes all black individuals.

air pollution to rise generally across the entire city, but left the Hill District relatively unimpaired because it was geographically less exposed to pollution from the steel mills.

To further explore the stability of pollution over time and the patterns in 1920, Table 4 shows the spatial correlation of pollution from different years. The table is constructed by interpolating each year's sootfall data to the centroids of the 1920 Enumeration Districts and then computing the correlation coefficients. There is clear persistence in the spatial patterns, but 1920 also stands out as an exceptional year. For example, looking at the bottom row, the correlation of the 1940 and 1920 data is much lower than the correlation between 1940 and the other years.

Based on the spatial interpolations, Appendix Tables A.1, A.2, and A.3 show average pollution levels as assigned to individuals in the full sample, linked sample, and county employee sample, respectively. Additionally, Table 5 shows the average pollution in each year for each of three demographic groups, black, foreign-born, and native-born white. Consistent with the patterns observed in modern times, in 1910, 1930, and 1940, native-born whites have lower pollution exposure on average than either blacks or foreign-born whites. However, consistent with the anomaly seen in Figure 5 and Table 4, this pattern is actually reversed in 1920, when blacks have lower pollution exposure and foreign- and native-born whites have approximately the same exposure.

4 Empirical Analysis

4.1 County Employee Sample

We begin with an evaluation of the relationship between income, demographics and air pollution in our Allegheny County employee data. These data have small sample sizes relative to the census data. However, they have the advantage of having actual incomes (vs. Occupational Income Scores). Furthermore, because they comprise individuals who all work for the same employer, they allow us to make better "apples to apples" comparisons among people.

At its core, the goal of this analysis is to identify how pollution exposure varied as a function of demographics. We do so by estimating the simple linear model presented in Equation 1:

$$\text{Std. Pollution}_i = \alpha + \beta \cdot ln[\text{income}]_i + \gamma \cdot \text{Black}_i + \delta \cdot \text{Foreign Born}_i + \varepsilon_i. \tag{1}$$

Here, Std. Pollution_i is the interpolated level of pollution either at the Enumeration District centroid where individual i lived (1910 and 1920 data) or at their exact address (1926 data). For each of our three time slices, pollution levels are normalized by the standard deviation of pollution across within-year observations. Income_i is individual i's annual income (aggregated up for hourly employees based on 40 hours/week and 50 weeks per year). Black_i and Foreign Born_i identify black and foreign-born employees respectively. Because of the small sample sizes, we focus on a limited set of demographic variables in these county employee regressions.

Results from these regressions are presented in Table 6. Columns 1-3 use the 1910 data. Column 1 regresses pollution on income, column 2 on indicators for black and foreign-born, and column 3 on all three variables. Columns 4-6 repeat this exercise with the 1920 data. Column 7 uses the data from 1926, when only income is available.

Focusing first on 1910, the point estimates indicate that a ten percent increase in income was associated with reduced levels of pollution exposure on the order of 0.35 standard deviations, an effect which is statistically significant. Comparing columns 1 and 3, we see that controlling for race and nativity does not affect this result. Looking at columns 2 and 3, the results indicate that on average both black and foreign-born residents bore a greater pollution burden in 1910, but that the effect is not statistically

Table 6: Regressing Pollution on County Employee Characteristics

	1910	1910	1910	1920	1920	1920	1926
Log Annual Income	-0.355***		-0.336**	0.0190		-0.0126	-0.142**
	(0.130)		(0.132)	(0.119)		(0.121)	(0.0640)
Black		0.399	0.372		-0.383**	-0.371*	
		(0.357)	(0.370)		(0.190)	(0.200)	
Foreign Born		0.0711	0.108		0.0646	0.0790	
		(0.199)	(0.210)		(0.142)	(0.150)	
Constant	6.283***	3.821***	6.116***	3.857***	4.027***	4.104***	5.618***
	(0.922)	(0.0484)	(0.937)	(0.877)	(0.0416)	(0.894)	(0.482)
N	226	463	226	437	659	437	1381
R^2	0.032	0.003	0.037	0.000	0.007	0.009	0.004
adj. R^2	0.028	-0.001	0.024	-0.002	0.004	0.002	0.003

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Pollution levels are expressed in terms of within study standard deviations.

significant. This lack of significance is, at least in part, driven by limited variation in our county samples.[10] Skipping 1920 for the moment, in 1926 income remains associated with reduced levels of pollution, with a ten percent increase in income implying a 0.14 standard deviation drop in pollution exposures, again statistically significant at conventional levels. There are no estimates on these variables for 1926 because, as discussed above, we could not match these individuals to census data.

Turning to 1920 (columns 4-6) we see a different result. The estimated income effect is essentially zero and being black is now associated with lower levels of pollution exposure. As discussed previously, this result is an artifact of a temporary reversal in pollution distribution that occurred around 1920 and that temporarily moved Pittsburgh's largest black neighborhood from near the top of the pollution distribution to near the bottom. The 1920 anomaly aside, these results suggest that as early as 1910 the types of income-driven Environmental Justice (EJ) patterns that we see today were already established and could be identified even when examining the residential locations of individuals who all worked for the same employer.

[10]The 1910 sample includes only 8 black employees and 27 foreign-born employees, while the 1920 sample includes 29 black employees and 54 foreign-born employees. Further, our analysis of city-wide census data (see below) finds that EJ results relative to being foreign-born are largely driven by individuals who do not speak English. Our county employee sample only includes one such foreign-born individual each in 1910 and in 1920.

4.2 Census Sample

We now turn to our main analysis based on data from the Decennial Census. Here, with a much larger sample, it becomes possible to consider a richer set of demographic characteristics and a broader range of years. However, it is now necessary to rely on Occupational Income Score as a proxy measure for income. For this analysis, we take as the unit of analysis household heads, as reported in the census. Thus, we are essentially combining outcomes for married couples (where the Census Bureau would have recorded the husband as the household head) with those for single men and women who are not boarders, housed in group quarters, nor living in the home of an older relative. As discussed in Section 3.2.1, 79% to 85% of household heads are male, depending on the year. We also consider sensitivity analyses with alternative samples.

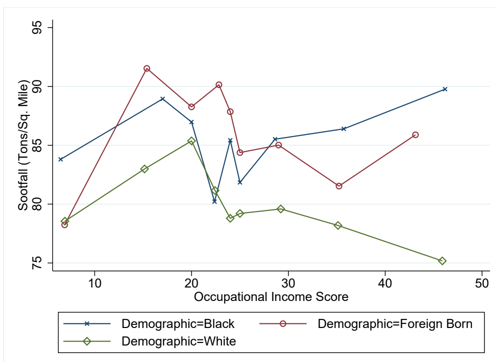
4.2.1 Patterns in the Raw Data

We begin by considering patterns in the raw data. As observed in Table 5, on average native-born whites were exposed to less pollution than black or foreign-born residents, except in 1920. To further explore these patterns and how they interact with income, we bin the data into deciles of the income distribution, then compute the average pollution experienced by each demographic group (native-born white, foreign-born, and black) in each decile. Figure 6 displays the results. It shows that, again excepting 1920, native-born whites experience lower pollution than black or foreign-born residents—throughout the income distribution. In 1910 and 1930, whites also appear to be sorting by income, with a clear downward trend in their exposure, but this pattern does not appear to be present for other groups.[11] This finding is consistent with the results for county employees presented in the previous sub-section, where we found exposure differences by income for that predominantly white sample. However, by 1940, we see a clear sorting pattern across demographic groups and across income within all three groups: at every decile of the income distribution, native-born white residents have the lowest exposure, foreign-born residents higher, and black residents higher still. Moreover, within each group, exposure falls with income. These patterns suggest that the sorting process was a dynamic one that unfolded between 1910 and 1940.

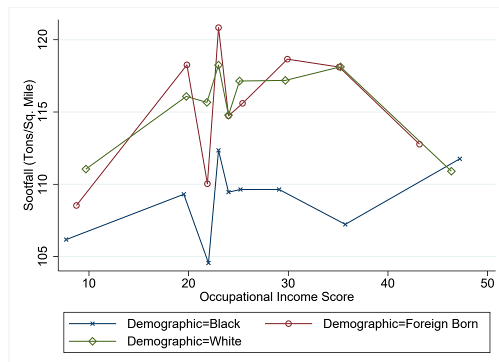
[11] Note the downward spikes in 1920 and 1930 at the 4th decile are for a mass in the distribution of Occupational Income Scores for low-level clerical workers such as telephone operators and typists, sandwiched between two groups of laborers.

Figure 6: Pollution vs. Income by Demographic Group (1910 thru 1940)

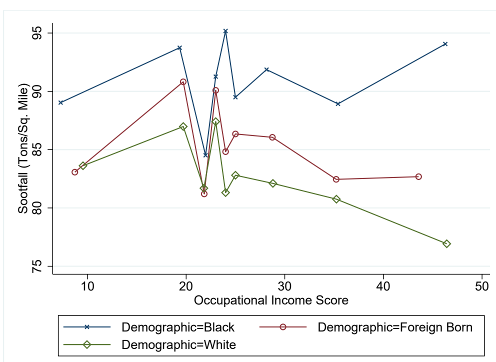
(a) 1910



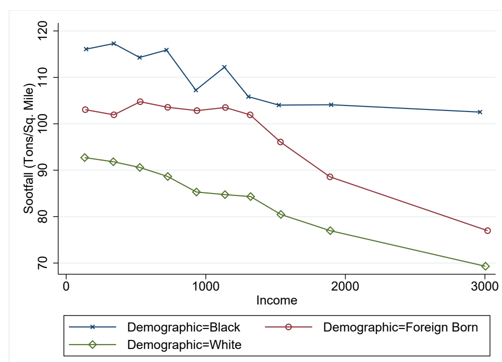
(b) 1920



(c) 1930

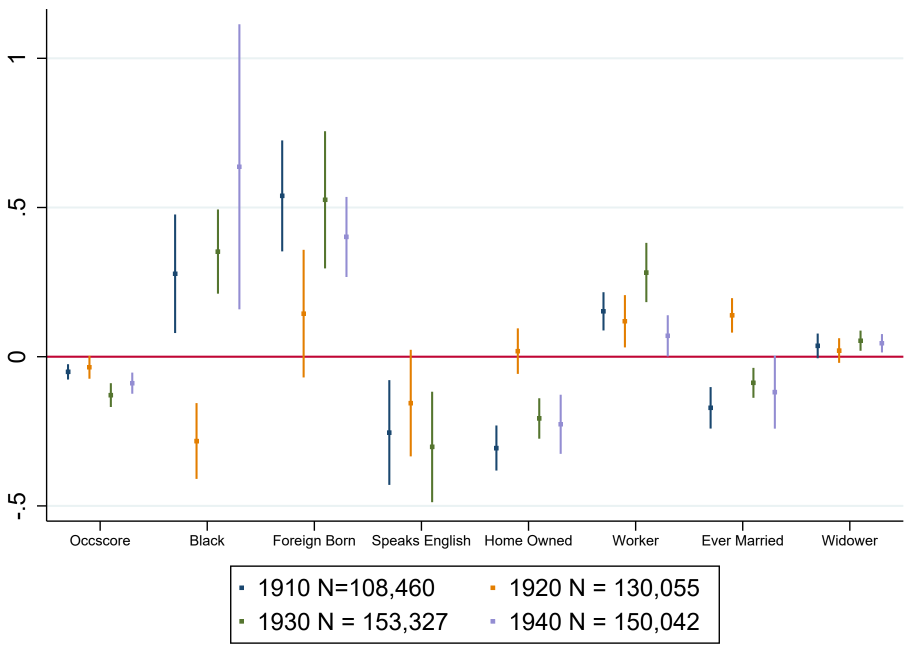


(d) 1940



This figure reports binned means of sootfall exposure, by demographics (White, Foreign Born, Black) where the bins are constructed based on either Occupational Income Score (1910, 1920 and 1930) or income (1940). Bin boundaries were chosen to equate populations across bins. For comparison, Appendix Figure A.1 reports the 1940 results based on Occupational Income Score.

Figure 7: Coefficient Plots from Cross-sectional Analysis with Full-Count Census



This figure reports a subset of the coefficients from estimating a version of Equation 1, expanded to include more demographic variables, on 4 separate samples (one for each decade). Household heads are the unit of observation. Pollution and income variables are standardized by within-sample standard deviations. For a complete set of results, see Appendix Table A.5.

4.2.2 Cross-Sectional Regressions

We further explore these patterns with multivariate regression, expanding the model of Equation 1 to look simultaneously not only at Occupational Income Scores, black, and foreign-born, but also English-speaking, home ownership, being in the labor force, and being widowed or divorced. We also control for sex, age, and an indicator for missing occupation. We estimate these regressions separately for each year. Although actual income is available in 1940, we use only Occupational Income Scores in this regression analysis for consistency with the earlier years. Figure 7 visually presents the results. For each variable, it displays the point estimate and 95% confidence intervals, separately for each year. All coefficients and standard errors are reported in Appendix Table A.5. In all models, standard errors are computed by clustering at the census division (Enumeration Districts for 1910-30 and Tracts in 1940).

In all four decades, we find that higher Occupational Income Scores for the household head are associated with lower levels of pollution exposure. In 1910 and 1920, ceteris paribus, a one standard deviation increase in proxied income is associated with pollution exposures that are reduced by between 4 and 5 percent of a standard deviation. For 1930 and 1940, the relationship is more pronounced, ranging between 9 and 13 percent.

Relative to income, with the exception of the anomaly in 1920, the results for race and nativity are much stronger. All else equal, black household heads in 1910 experience pollution levels that are 28% higher than their white counterparts. By 1940, the disparity increases to 64%. For foreign-born household heads, the penalty is relatively stable, ranging between 40 and 54 percent. Assimilation (speaking English) appears to offset roughly half of this penalty.[12]

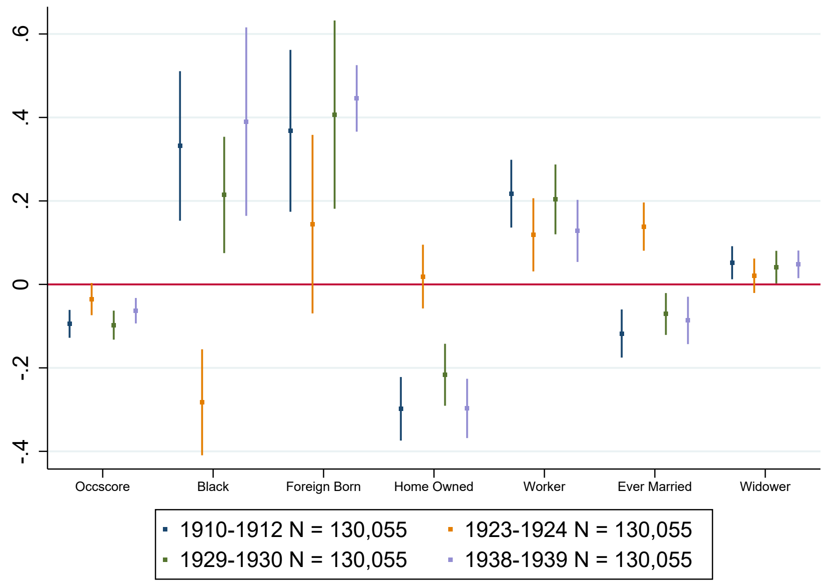
Continuing to focus on 1910, 1930 and 1940, home ownership is associated with lower levels of pollution exposure as is being married. Conversely, being in the work force and/or being a widow or widower is associated with higher levels of pollution exposure. Overall these estimates suggest there was meaningful variation in pollution exposure across a range demographic characteristics. Perhaps most striking is how much larger is the estimated impact of race and nativity relative to these other characteristics, particularly given that all demographic variables are included in a single regression so these are ceteris paribus comparisons. And, as we explore further below, for black Pittsburghers the relationship appears to strengthen significantly over our sample period—30 years that overlap the initial wave of the Great Migration and saw a rapid increase in racial segregation in the city.

Our conclusions are not sensitive to the sample definition using household heads. Appendix Tables A.6 and A.7 report results using only single men and single women respectively. These samples include boarders and individuals living with family members. In Subsection 4.3 below, we also discuss results with a full sample. In all cases, Occupational Income Score is negatively associated with pollution exposure, being foreign-born is positively associated with pollution, and, except for 1920, being black is also positively associated with pollution.

On the other hand, our results are more sensitive to year, with insignificant and sometimes counter-intuitive results in 1920. We have argued that, particularly for black individuals and their families, the 1920 pollution distribution is a relatively short-lived anomaly. To test this interpretation more directly, we replicate our analysis for household heads, but hold demographics constant at their 1920 levels. That is, we regress each

[12] The Census didn't ask about speaking English in 1940.

Figure 8: Exploring the 1920 Pollution Anomaly



This figure reports results for four separate regressions. We fix demographics at their 1920 levels for all four regressions and only vary the pollution measures across these regressions. Thus there is one regression for each decade's pollution levels (1910, 1920, 1930 and 1940.). Pollution and income variables are standardized by within sample standard deviations. For a complete set of results, see Appendix Table A.9.

year's pollution on 1920 demographics. Mechanically, the results from these regressions will be the same as those reported above in Figure 7 for 1920. However, for the other years they will be different. The idea is to test whether, in 1920, population was still sorted on long-run pollution averages, if not on actual 1920 pollution.

Figure 8 illustrates the results. (See Appendix Table A.9 for a complete report of the underlying statistics.) The figure shows that for all periods, except 1920, pollution levels were elevated in the neighborhoods where black people were concentrated in 1920. They are also higher in areas with more foreign-born residents, and lower in areas with higher income and higher home ownership. In other words, the reversed relationships we see in 1920 appear to be associated with a short-run change in the spatial distribution of pollution rather than a re-sorting of individuals across neighborhoods vis-a-vis pollution.

To better understand the relationship between these results and the great migration, in Table 7 we reproduce the analysis of Figure 7 that includes controls for being southern born and interacting the southern born indicator with an indicator of being Black. For

Table 7: Pollution Exposure HH Heads w/ Southern Migrants

	Census 1910	Census 1920	Census 1930	Census 1940
Southern Born	-0.0399	-0.211***	-0.117***	0.120*
	(0.0405)	(0.0568)	(0.0308)	(0.0657)
Black	0.430***	-0.225***	0.348***	0.602**
	(0.105)	(0.0684)	(0.0742)	(0.274)
Southern Born X Black	-0.201***	0.117*	0.130***	-0.0602
	(0.0623)	(0.0656)	(0.0421)	(0.0946)
Observations	61679	81481	105552	112939
R^2	0.042	0.011	0.047	0.085
Adjusted R^2	0.042	0.011	0.047	0.085

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Pollution levels are expressed in terms of within census standard deviations. Foreign born individuals are dropped. SEs clustered at the Census division.

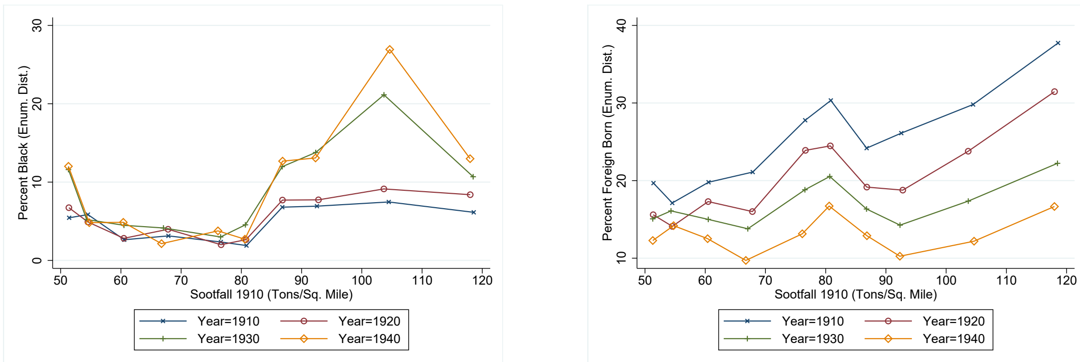
an apples to apples comparison, we exclude foreign born individuals from this analysis. The broad pattern of results is unchanged, so we only report the coefficients of immediate interest here (complete results are reported in Appendix Table A.8). In 1910, at the beginning of the Great Migration, Black Southern Migrants are experiencing lower pollution levels than those who were not born in the South. Whereas, for whites born in the south, there appears to be no difference. As of 1930, the relationship has reversed. Here, white southern migrants are experiencing lower pollution levels than their brethren, while Black southern migrants are experiencing the same elevated levels of pollution as their northern-born counterparts.

The southern-born population in Pittsburgh (both Black and White) grew roughly 10% per decade between 1910 and 1930. Between 1930 and 1940, this trend reversed - with southern-born populations shrinking by more than 10% for both Black and White Pittsburghers. By the end of this decade, both Black and White southern-born migrants were, on average, exposed to roughly a tenth of a standard deviation more pollution than their northern-born counterparts.

Figure 9: Demographic Sorting Relative to 1910 Sootfall

(a) Percent Black

(b) Percent Foreign Born



Note: This figure reports the mean Black percentage and mean Foreign Born percentage as a function of 1910 sootfall for the years 1910 through 1940. Bin-widths are chosen such that each bin has the same population.

4.2.3 Dynamic Sorting on Pollution

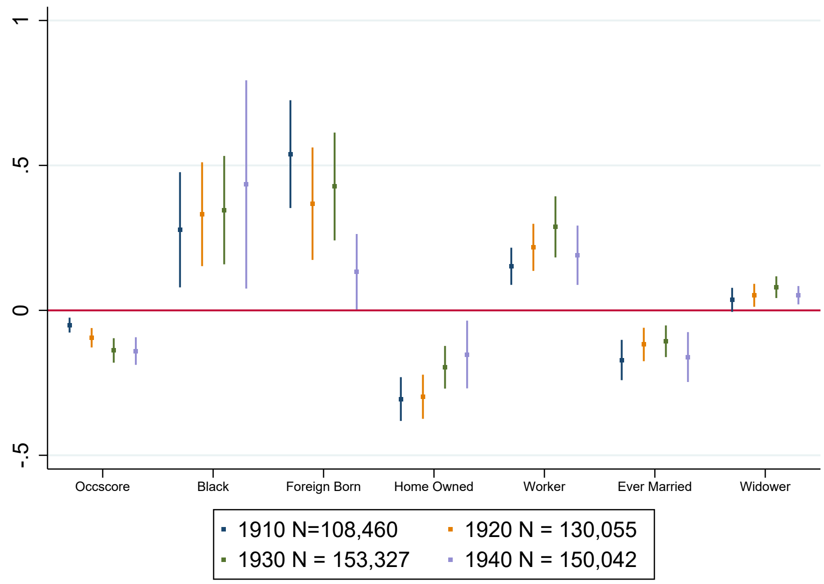
In a somewhat similar vein, we can also look across decades to shed light on the dynamics of pollution exposure. One key finding in our analysis so far has been the fact that, as the Great Migration increased the size of Pittsburgh's black population and the level of racial segregation increased, the correlation between race and pollution grew substantially. In general, the spatial distribution of pollution sources, and hence pollution exposure, was constrained by the basic geography of the city and the heavy reliance of its industrial base on access to the rivers.

Thus, this increased correlation was likely the result of the sorting of individuals into neighborhoods rather than the systematic location of pollution sources into black neighborhoods. However, it is important to emphasize that this "sorting of individuals" was occurring in a highly discriminatory environment rife with formal and informal racism, particularly in housing markets, and thus shouldn't be interpreted as representing systematic differences in preferences for air quality (Boustan 2010; Rothstein 2017; Shertzer and Walsh 2019).[13]

As a starting point for considering sorting in response to pollution, we create an index of how demographics have responded to lagged pollution. Specifically, for each

[13] Christensen and Timmins (2022, 2023) consider the relative effects of "steering" and voluntary "sorting" in today's markets, as well as the effect of discriminatory constraints on the housing process.

Figure 10: Dynamic Sorting of Household Heads (Dep. Var. = 1910 Sootfall)



Note: Note: We report results for four separate regressions. We fix pollution at 1910 levels for all four regressions and only vary the demographic measures across these regressions. Thus there is one regression for each decade's demographics(1910, 1920, 1930 and 1940.). Pollution and income variables are standardized by within sample standard deviations. For a complete set of results, see Appendix Table A.10.

census, we attach 1910 pollution, order the data by 1910 pollution exposure, cut it into ten bins, and calculate the percent black, percent foreign-born, and the average sootfall within each bin. Figure 9 presents the results. Looking at Panel a, we see that areas that were initially polluted in 1910 had higher black populations in 1910, as previously noted, but now we also see that in those same areas the black population grew over time relative to other areas. That is, as the Great Migration proceeded, black households were systematically moving into the areas that were initially polluted. The fact that areas with greater 1910 pollution had more black residents in 1920 also corroborates our findings in Figures 7 and 8 that 1920 exposure was an anomaly, with black residents still sorted by lagged pollution. In Panel b, we see the opposite pattern holds for foreign-born. The most polluted areas had very high concentrations of foreign-born residents in 1910, but each subsequent decade saw a drop in their share in those areas.

To further explore how demographics adjusted over the next 30 years relative to these "initial" conditions, we regress 1910 sootfall levels on demographics in each of the four Census waves. Figure 10 displays the results and Appendix Table A.10 provides the underlying coefficients and standard errors. Confirming the trends in the raw data seen in Figure 9, Figure 10 shows that between 1910 and 1940, as the city became more racially segregated, black residents became increasingly concentrated in those areas that had been heavily polluted in 1910. This result reinforces our finding from the repeated cross-sections that the correlation between race and pollution increased over this time period and suggests that this increase was, at least in part, due to the sorting of individuals. The relationship between income and pollution also strengthens over this period. The opposite is true for nativity, where we see that over our 30 year period immigrants became less concentrated in those areas that were most polluted in 1910.

4.3 Panel Data of Movers and Stayers

The analysis of the previous sub-section is based on repeated cross sections. However, the 1940 Census reports where residents lived five years prior, in 1935. Additionally, the Census Tree Project tracks a panel of individuals and their locations across Census waves (Price et al. 2021). These data allow us to analyze mobility patterns on an individual level.

We begin by focusing on race and nativity. We estimate the following model.

$$\begin{aligned}\text{Std. Pollution}_{i,1940} &= \alpha + \beta \cdot \text{Black}_i + \gamma \cdot \text{Foreign Born}_i + \delta \cdot \text{Change Nbrhd}_i \\ &+ \theta \cdot \text{Black X Chg Nbrhd}_i + \phi \cdot \text{For. Born X Chg Nbrhd}_i \\ &+ \mu' X_i + f(\text{income}_{i,1940}, \text{occ score}_{i,\text{lag}}) + \varepsilon_i\end{aligned} \tag{2}$$

That is, we regress pollution exposure in 1940 (expressed in standard deviations) on race (black/white), an indicator for whether the individual moved from one Pittsburgh neighborhood to another during the relevant time period (with staying being the alternative), and an interaction between the two. We do the same for nativity.[14] We also include either controls for 1940 income (specifications 1 and 2) or flexible controls for the pattern and levels of occupational income scores in both periods (columns 3 thru 6), as well as a full suite of 1940 demographic variables, X.[15]

[14] Because we only observe pollution in Pittsburgh, in all models, every individual included in the sample was living in Pittsburgh in 1940. Furthermore, they were in the region in both time periods—in Allegheny County in 1935 or in Pittsburgh in 1930 or 1910.

[15] The analyses in specifications 1 and 2 are based solely on the data from the 1940 census and, in the case of specification 2, its question about retrospective location 5 years earlier. As a result, we only observe demographic data in 1940 for this specification and can therefore use actual income in our estimation. In specifications 3 thru 6 we are linking across censuses and thus have income score measures in two periods but only have actual income for the second period. Here, we switch to using Occupational

Table 8: Sorting by Race and Pollution

	1940 Full	1940 Full	1940 - 1930	1940 - 1910	1940 - 1930	1940 - 1910
Black	0.553**	0.271*	0.190	0.0268	-1.285	-5.084
	(0.217)	(0.141)	(0.130)	(0.157)	(3.685)	(4.313)
Foreign Born	0.366***	0.346***	0.366***	0.350***	6.120***	4.402**
	(0.0623)	(0.0568)	(0.0600)	(0.0593)	(1.683)	(1.716)
Change Nbrhd		-0.100***	-0.202***	-0.362***	-8.856***	-17.43***
		(0.0284)	(0.0382)	(0.0653)	(2.474)	(2.547)
Black X Chg Nbrhd		0.418***	0.314***	0.455***	10.23***	18.10***
		(0.130)	(0.0909)	(0.120)	(3.286)	(4.010)
FB X Chg Nbrhd		0.0100	-0.0543	-0.0646	-4.469***	-7.963***
		(0.0275)	(0.0438)	(0.0556)	(1.595)	(2.145)
Income 1940	-0.0950***	-0.0909***				
	(0.0157)	(0.0150)				
Dep. Var.	1940 Soot	1940 Soot	1940 Soot	1940 Soot	Chg Soot Pctile	Chg Soot Pctile
Flex. Inc. Cntrls			X	X	X	X
N	415505	415505	297565	85363	297565	85363

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Pollution levels and incomes are expressed in terms of within census standard deviations. All regressions include a complete set of 1940 controls (see Appendix Table A.11 for a complete list of coefficient estimates). Columns 1 and 2 include all individuals in the 1940 census, with changes in neighborhood and/or county being based on the census' 5 year move move variable. Columns 3 and 4 match samples of men from the 1940 census to the 1930 and 1910 censuses respectively. Columns 5 and 6 use the matched samples as well, with the dependent variable being the change in pollution percentile from period 1 to period 2.

Table 8 presents the results for the key coefficients of interest. (See Appendix Table A.11 for the full set of coefficient estimates.) Columns 1-4 consider how mobility affects final exposure in 1940, and so use 1940 sootfall as the dependent variable. Columns 1 and 2 include all individuals in the 1940 Census—not just household heads—who reported living in Pittsburgh in 1940 and in Pittsburgh or the surrounding Allegheny County in 1935. All of the samples used in this table pool single men, single women and all married individuals. Because it differs from the Household Head samples used above, in column 1 we first replicate our earlier cross-sectional analysis. The results mirror those of the earlier tables. The relative importance of race and income re-inforce results found in the earlier analysis—particularly demonstrating that our results on income are not sensitive to the use of actual income as opposed to Occupational Income Score (at least for 1940). In column 2, we now include an indicator for having

Income Score and control flexibly for different patterns of labor force participation, missing Scores and levels of the Scores in each period.

changed neighborhoods over the last five years, as well as interactions with black and foreign-born. Native-born white non-movers are the omitted category. Thus, the coefficient on black can be interpreted as the differential exposure (relative to non-moving white residents) for non-moving black residents. We find that these individuals were, on average, exposed to 0.27 standard deviations more pollution in 1940 than were their non-moving white counterparts. For non-moving foreign-born residents, the difference is 0.35 standard deviations. The marked importance of racial differences in neighborhood sorting is demonstrated by the coefficients on the interaction terms between race and changed neighborhood. Native-born whites who reported moving over the previous 5 years were exposed to roughly a tenth of a standard deviation *less* pollution than non-moving whites. The interaction for foreign-born is small and statistically insignificant, indicating that their mobility patterns were comparable to native-born whites. In contrast, black residents who reported moving in the last 5 years were exposed to 0.32 standard deviations *more* soot than were non-moving blacks (0.418-0.100). Thus, mobility widened the pollution gap: the coefficient on the interaction term indicates that differential patterns between white movers and black movers yielded an estimated movement-generated gap of an additional 0.42 standard deviations.

Columns 3 and 4 include samples of inidividuals who were successfully linked across censuses by the Census Tree Project and who report living in Pittsburgh in both periods. Column 3 uses matched census samples between 1930 and 1940 and column 4 uses 1910 to 1940.[16] As discussed in Section 3.2.2, the requirements of the matching process leave us with a smaller linked sample. Because we observe occupations in 1930 and 1910, we can control flexibly for both baseline and final income as well as changes in income. In column 3, looking back over a 10 year window with our matched sample yields fairly similar results to those of column 2. Over the 10-year window, Black movers were associated with 0.31 standard deviations more sootfall than non-Black native White movers, vs. 0.42 for the 5-year window. Finally, column 4 looks back to 1910. Over this 30 year period, the movement-generated gap between black and white movers is estimated to be 0.46 standard deviations. For foreign-born, moves result in less pollution, at a rate roughly comparable to native-born whites over the entire period, but even faster than native-born whites over the decade of the 1930s.

While the first four models focus solely on final pollution exposure at the end of the

[16] In these columns we define an individual as having changed neighborhoods if their 1940 census division does not intersect with their census division from the prior period.

period (1940), we can also explore *changes* in exposure. Continuing to use the linked data, we thus regress 1940-1930 and 1940-1910 changes in sootfall exposure, respectively, on our interactions of race and mobility. Columns 5 and 6 of Table 8 present the results of this analysis. Here, the dependent variable is the change in an individual's soot exposure percentile—i.e., soot exposure percentile in 1940 minus soot exposure percentile in either 1930 or 1910. For the 1930 to 1940 period we estimate a movement-generated gap of roughly 9 percentile points for black men, while for 1910 to 1940 we estimate a movement-generated gap of roughly 18 percentile points. For foreign-born men, we estimate that they are moving to cleaner areas even faster than native-born whites. This result is consistent with the finding of Figure 10, where we see a notable drop in the partial correlation between pollution and foreign-born in 1940.

While the regressions in Table 8 control for income, there is still a concern that the reported results could somehow reflect inherent differences in income or wealth between black and white Pittsburghers at the time. To explore this possibility, we split the sample in half at the median income and re-estimate the models from columns 2 and 3 of Table 8.[17]

Table 9 reports the results. Columns 1 and 2 are based on the 1935-40 mobility question from the full-count 1940 Census. Because we don't observe incomes for 1935, this model splits the sample at the median income for 1940, which was $1,190. Columns 3 and 4 use the linked data between the 1930 and 1940 censuses. Because we prefer to use lagged income where possible, here we split the sample by 1930 Occupational Income Scores (at a median score of 24). For both samples, the results are stable across the income split.

Collectively, the results of Figures 9 and 10 and Tables 8 and 9 demonstrate that systematic differences in the racial sorting of individuals and the associated increase in racial segregation that occurred in Pittsburgh over this period led to marked increases in racial inequity in exposure to pollution. In 1910, white foreign-born residents vastly outnumbered black residents, but both were segregated into higher polluted areas. Over the ensuing years of the Great Migration, both native-born whites and foreign-born residents systematically moved to less polluted areas, while black residents moved to more

[17]Note that we omit an analysis of the 1910-1940 linked sample due to the small number of black men earning above the median income in 1910; for this linkage there are less than 100 black men in the sample. For the 1930-1940 linkage, we have 1,227 black men below and 252 black men above the median Occupational Income Score. In the 1935-1940 sample, we have 8,201 black men below the 1940 median income and 2,676 black men above the 1940 median income.

Table 9: Sorting by Race and Pollution: Income Split

	Below Median	Above Median	Below Median	Above Median
Black	0.231	0.275**	0.227	0.233
	(0.148)	(0.107)	(0.150)	(0.158)
Foreign Born	0.331***	0.289***	0.353***	0.323***
	(0.0603)	(0.0626)	(0.0708)	(0.0603)
Changed Neighborhood	-0.0865***	-0.0751***	-0.187***	-0.174***
	(0.0302)	(0.0203)	(0.0472)	(0.0449)
Black X Neighborhood	0.361***	0.309***	0.297***	0.422**
	(0.118)	(0.106)	(0.0880)	(0.164)
FB X Chg Nbrhd	0.0349	-0.00523	-0.140*	-0.141**
	(0.0312)	(0.0260)	(0.0728)	(0.0613)
Annual Income	-0.104***	-0.0997***		
	(0.0241)	(0.0162)		
Dep. Var.	1935-1940	1935-1940	1930-1940	1930-1940
Flex. Inc. Cntrls			X	X
N	64426	64351	10589	9796

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Pollution levels are expressed in terms of within census standard deviations. For the 1935-1940 the sample is split above and below $1,190 in 1940 income. For the 1930-1940 the sample is split above and below a 1930 Occupational Income Score of 24.

polluted areas. These patterns resulted in a notably widened black-white difference by 1940.

4.4 The Role of Prices

We conclude our empirical analysis with an assessment of the salience of sootfall in Pittsburgh's housing market over this period. If pollution was an aspect of housing location that was relevant to people's location choices, it should have been capitalized into housing prices. We explore this issue here.

The Census first asks about home values and rents in 1930. Thus, our systematic analysis of the pollution-price relationship is limited to 1930 and 1940. However, even for the earlier period, there is anecdotal evidence supporting a link. As noted previously, in his study of smoke's damages in Pittsburgh, O'Connor (1913) surveyed local real estate experts about the effect of soot on property values and rents. While a handful of responses argued that impacts were negligible or impossible to assess, the majority discussed extreme examples where property values/rents were discounted on the order

Table 10: Hedonic Analysis (Dep. Var. = Housing Price/Rent

	1930	1930	1940	1940	1930	1940
Tons/Sq Mile Soot	-0.0587***	-0.0547***	-0.0512***	-0.0487***	0.00502	0.0463
	(0.00688)	(0.0137)	(0.00641)	(0.00777)	(0.0298)	(0.0314)
Dep. Variable	Price 1930	Rent 1930	Price 1940	Rent 1940	Price 2010-23	Price 2010-23
Observations	12635	3773	11345	6296	6369	5665
Adjusted R-squared	0.658	0.416	0.601	0.540	0.415	0.374

Standard errors in parentheses, clustered by Census division.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Pollution levels are expressed in terms of within census standard deviations.

of 25 to 50 percent. A typical response was that of real estate agent Edward Lang who stated, "All the property in this section (that is residence property) has depreciated fully 50% by reason of the excessive smoke nuisance."[18]

For a more systemic analysis, we utilize the linked hedonic data set for 1930 and 1940 discussed in Section 3.3. We estimate the model presented in Equation 3 separately for owner occupied and rental units in each of 1930 and 1940,

$$lnP_j = \alpha + \beta \cdot \text{Sootfall}_j + \Gamma' Z_j + \varepsilon_i \tag{3}$$

where P_j is either housing value or rent, respectively, in neighborhood j and Sootfall_j is normalized by its standard deviation. The vector Z_j represents housing characteristics, including indicator variables for # of bedrooms, # of bathrooms and total rooms interacted with total living space, as well as age and age^2, an indicator for new build, lot area and a set of neighborhood indicators. We trim for outliers at the one percent level based on rent, price, lot size and square feet of living space. When modeling sales price, we also trim at the 95% for the ratio of 2022 assessed value to 1930/1940 reported value in an attempt to further tease out any potentially inaccurate matches. Our results are not sensitive to this trimming.

The results from this analysis are quite robust and are presented in the first four columns of Table 10. Columns 1 and 2 present results for 1930, for asset values and rents respectively. Columns 3 and 4 do the same for 1940. On average, a one standard deviation increase in sootfall is associated with price and/or rent discounts roughly on the order of 5 to 6 percent.

[18]For more on the broader activities of the Mellon Institute see Banzhaf and Walsh (2024). The specific letter quoted here is located in the University of Pittsburgh Archives, AIS.1983.07.Ser.1.ff8.

Table 11: Household Head Regressions with Price Controls

	1930	1940
Occscore	-0.0847***	-0.0736***
	(0.0153)	(0.0161)
Black	0.328***	0.628**
	(0.0714)	(0.243)
Foreign Born	0.451***	0.394***
	(0.115)	(0.0677)
Speaks English	-0.253***	
	(0.0936)	
Home Owned	-0.331***	-0.273***
	(0.0488)	(0.0631)
Worker	0.193***	0.0473
	(0.0395)	(0.0325)
Widower	0.0378**	0.0431***
	(0.0165)	(0.0150)
Divorced	-0.00714	-0.0302
	(0.0340)	(0.0390)
Female	0.0248	-0.0930***
	(0.0190)	(0.0244)
Age	0.00208***	-0.00262***
	(0.000627)	(0.000789)
Missing Occ.	-0.237***	-0.0691
	(0.0433)	(0.0659)
Ever Married	-0.0985***	-0.124**
	(0.0248)	(0.0623)
Rent	-0.00416***	-0.00247***
	(0.000797)	(0.000568)
Home Value	-0.00650***	-0.00752**
	(0.00148)	(0.00292)
Constant	3.923***	2.049***
	(0.0663)	(0.132)
Observations	148261	146296
R^2	0.060	0.086
Adjusted R^2	0.060	0.086

Standard errors in parentheses, clustered by Census division.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The Table reports the coefficients from estimating a version of Equation 1, expanded to include more demographic variables, on 2 separate samples, one for 1930 and one for 1940. Household heads are taken as the unit of observation. Pollution and income variables are standardized by within sample standard deviations. Rents are monthly and home value is measured in thousands of dollars.

Earlier, we argued that pollution and demographic compositions were not simultaneously determined, as the location of industrial facilities was largely dictated by geography. However, this raises the potential concern that these price effects are being driven by that geography, e.g. distance to the river, distance to the central business district, or elevation. But if such persistent factors, potentially correlated with pollution, were driving our results, we would expect them to continue to affect prices today. Accordingly, as a placebo test, in columns 5 and 6 we regress *recent* sales prices (from 2010-2023) on 1930 and 1940 pollution levels respectively. To address price appreciation over the 13 year period between 2010 and 2023, we add year fixed effects to the model of Equation 3. These last results show that, for modern day Pittsburgh, when sootfall is no longer a daily concern, the spatial pattern of prices no longer has a relationship with the patterns of pollution from 70+ years ago. This finding suggests that our pollution results are not driven by some correlation between sootfall and other underlying and persistent geographic patterns of location preference.

Given these findings on housing prices, it is logical to ask if the racial patterns that we measure here are being driven solely by black households choosing to live in cheaper housing, which is correlated with higher pollution levels. While we control for income differences in these regressions, it is well documented that, conditional on income, black wealth levels are systematically much lower than are white wealth levels (Derenoncourt et al. 2023). Thus, even with income controls there could be scope for a direct effect of housing prices to drive our results. To test this channel, in Table 11 we replicate our baseline specification adding housing prices and rents to the set of controls. Consistent with the hedonic regression results, households living in housing with higher rents or home values are, *ceteris paribus*, exposed to cleaner air. Moving from the 25th to 75th percentile in rent/value is associated with lower air pollution exposures on the order of 0.03 to 0.15 standard deviations. However, inclusion of these controls only leads to very modest reductions in the estimated black coefficient (a 6.8% reduction for 1930 and a 1.1% reduction for 1940).

These final results provide clear evidence that air pollution (sootfall) was salient for Pittsburghers over the time frame we study—a result that is not surprising given the anecdotal evidence from the day. However, while pollution was salient in housing markets, our analysis also shows that the dynamic patterns of racial differences in exposure to pollution which we observe were almost certainly not just a by-product of black households choosing to buy/rent cheaper homes in general.

5 Conclusion

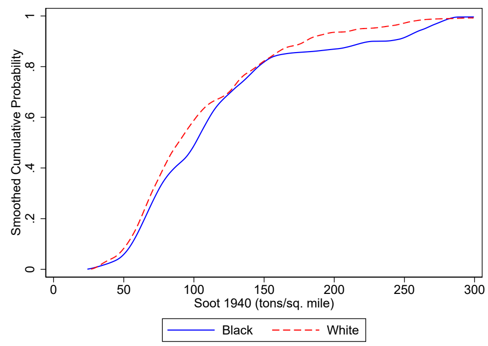
We explore the relationship between air pollution exposure and demographic characteristics in Pittsburgh during the period of 1910 to 1940, which witnessed a marked rise in racial segregation in Northern urban centers. Utilizing a unique and recently digitized historical panel data set on the spatial distribution of air pollution in Pittsburgh, perhaps the most polluted city in the pre-war United States, our findings reveal that race and nativity were much stronger predictors of exposure to pollution than income. Specifically, black Pittsburgh residents were exposed to significantly higher levels of pollution than their white counterparts with similar demographic characteristics. This disparity grew to roughly three quarters of a standard deviation in pollution exposure by 1940.

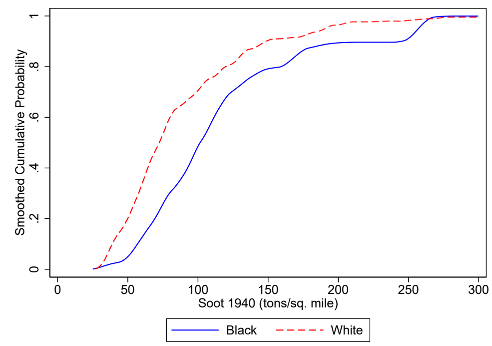
Our analysis further reveals that within-city moves were a key driving force behind the increasing correlations between race and pollution exposure. Relative to white non-movers, white movers experienced lower levels of pollution. However, for black residents, this relationship was reversed, with black movers being exposed to higher pollution levels than non-moving black residents. Taken together, these results suggest that the cumulative effect of moving on black-white exposure gaps was between one half to three quarters of a standard deviation in contemporaneous pollution exposure. This dynamic is likely due to a combination of income and wealth disparities and formal and informal discrimination (Boustan 2010; Rothstein 2017; Shertzer and Walsh 2019).

Meanwhile, foreign-born whites appear to have been assimilating with their native-born counterparts. While in 1910 they were segregated into more polluted areas just like black residents, over time they moved to cleaner areas at the same rate as native-born whites or even faster. This pattern is consistent with the finding of Abramitzky et al. (2021) that immigrants have had high rates of upward mobility throughout US history. By 1940, black-white pollution gaps were much more stark than gaps by nativity.

We also shed light on the salience of air pollution in Pittsburgh's housing market, specifically the relationship between sootfall and housing prices in 1930 and 1940. Our analysis finds that a one standard deviation increase in sootfall was associated with both sales price and rental discounts of between 5 and 6 percent, providing perhaps the earliest evidence to date of the capitalization of air pollution into housing prices. Further analysis demonstrates that these systematic differences in housing prices are not driving our main results on race, nativity and income.

Figure 11: Summarizing Exposure to 1940 Sootfall





Finally, to summarize the process of Pittsburgh's increasing racial segregation and the growth of black-white pollution gaps, Figure 11 summarizes how dramatically increased levels of racial segregation between 1910 and 1940 were associated with differential changes in pollution exposure. Panel (a) presents what the Cumulative Density Function (CDF) of exposure to sootfall would have been for black and white residents in 1940 if demographic patterns had remained fixed at their 1910 levels, while Panel (b) reports the distribution that was actually experienced as a result of demographic sorting and segregation. While gaps would still have existed in 1940 had racial segregation not grown, they would have been substantially lower than what actually occurred. For example, in Panel (a) we see that at the 50th percentile without any demographic sorting, white individuals were already destined to be exposed to 12 percent less sootfall than were black households. However as we see in Panel (b) demographic sorting associated with increased racial segregation resulted in a realized gap at the 50th percentile of 30 percent. At the 90th percentile, the sorting related growth in this gap was even more pronounced, growing from 21 percent to 40 percent.

Thus, the host of individual and institutionalized factors that drove changing demographic patterns in Pittsburgh over this time period not only led to a more segregated city, but to one where black households were disproportionately concentrated in polluted neighborhoods. Further, given that in 1910 a significant number of black households were living in neighborhoods that were destined to be at relatively low pollution levels in 1940, racial segregation didn't need to lead to this increased pollution exposure gap. In contrast, although pollution gaps persist, the magnitudes are falling, but less from resorting than from environmental improvements in high-pollution areas (Currie et al. 2023). Thus, the historical establishment of black-white pollutions gaps was rooted in different socio-economic mechanisms than their present-day unwinding.

Our findings have important implications for the persistence of racial and economic inequalities in the United States. Taken together, they suggest that patterns of racial and economic inequality in Pittsburgh were likely reinforced by disparate exposure to air pollution. This concern is particularly significant when considering the growing evidence of the link between pollution exposure and a broad range of developmental, material, and health outcomes. While the pollution experienced in Pittsburgh was particularly acute, pollution levels were generally elevated in industrial cities across the U.S. during this time period. Given the broad prevalence of racialized housing markets across virtually all U.S. cities throughout this era, these results likely have implications that extend beyond the City of Pittsburgh.

References

Abramitzky, R. and L. Boustan (2017). Immigration in American economic history. *Journal of Economic Literature 55*(4), 1311–1345.

Abramitzky, R., L. Boustan, E. Jácome, and S. Pérez (2021). Intergenerational mobility of immigrants in the United States over two centuries. *American Economic Review 111*(2), 560–608.

Akbar, P., S. Hickly, A. Shertzer, and R. Walsh (2023). Racial segregation in housing markets and the erosion of black wealth. *Review of Economics and Statistics*.

Almond, D., J. Currie, and V. Duque (2018). Childhood circumstances and adult outcomes: Act II. *Journal of Economic Literature 56*, 1360–1446.

Banzhaf, H. S., W. H. Desvousges, and F. R. Johnson (1996). Assessing the externalities of electricity generation in the Midwest. *Resource and Energy Economics 18*(4), 395–421.

Banzhaf, H. S. and O. Farooque (2013). Interjurisdictional housing prices and spatial amenities: Which measures of housing prices reflect local public goods? *Regional Science and Urban Economics 43*(4), 635–648.

Banzhaf, H. S., L. Ma, and C. Timmins (2019). Environmental justice: The economics of race, place, and pollution. *Journal of Economic Perspectives 33*(1), 185–208.

Banzhaf, H. S. and R. Walsh (2008). Do people vote with their feet? An empirical test of Tiebout's mechanism. *American Economic Review 98*(3), 843–863.

Banzhaf, H. S. and R. Walsh (2024, April). Smoke from factory chimneys: The applied economics of air pollution in the Progressive Era. *National Bureau of Economic Research WP 32328*.

Beach, B. and W. Hanlon (2018). Coal smoke and mortality in an early industrial economy. *Economic Journal 128*, 2652–2675.

Benner, R. (Ed.) (1913). *Papers on the Effect of Smoke on Building Materials. Bulletin No. 6. Mellon Institute of Industrial Research and School of Specific Industries*. Pittsburgh: University of Pittsburgh.

Bishop, K. C., J. D. Ketcham, and N. V. Kuminoff (2023). Hazed and confused: The effect of air pollution on dementia. *Review of Economic Studies 90*(5), 2188–2214.

Bondy, M.and Roth, S. and L. Sager (2020). Crime is in the air: The contemporaneous relationship between air pollution and crime. *Journal of the Association of Environmental and Resource Economists 7*(3), 555–585.

Borgschulte, M., D. Molitor, and E. Zou (2022). Air pollution and the labor market: Evidence from wildfire smoke. *National Bureau of Economic Research Working Paper w29952*.

Boustan, L. P. (2010). Was postwar suburbanization "white flight"? Evidence from the black migration. *The Quarterly Journal of Economics 125*(1), 417–443.

Brimblecombe, P. (1987). *The Big Smoke: A History of Air Pollution in London Since Medieval Times*. London: Methuen.

Burkhardt, J., J. Bayham, A. Wilson, B. J. D. Carter, E., K. O'Dell, ..., and J. Pierce (2019). The effect of pollution on crime: Evidence from data on particulate matter and ozone. *Journal of Environmental Economics and Management 98*.

Chen, Y., A. Ebenstein, M. Greenstone, and H. Li (2013). Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. *Proceedings of the National Academy of Sciences 110*(32), 12936–12941.

Christensen, P. and C. Timmins (2022). Sorting or steering: The effects of housing discrimination on neighborhood choice. *Journal of Political Economy 133*(8), 2110–2163.

Christensen, P. and C. Timmins (2023). The damages and distortions from discrimination in the rental housing market. *Quarterly Journal of Economics 138*(4), 2505–2557.

Clevenger, J. (1913). *The Effect of the Soot in Smoke on Vegetation. Bulletin No. 7. Mellon Institute of Industrial Research and School of Specific Industries*. Pittsburgh: University of Pittsburgh.

Collins, W. and R. Margo (2006). Historical perspectives on racial differences in schooling in the United States. *Handbook of the Economics of Education 1*, 107–154.

Collins, W. and R. Margo (2011). Race and home ownership from the end of the civil war to the present. *American Economic Review 101*(3), 355–359.

Colmer, J. and J. Voorheis (2020). The grandkids aren't alright: The intergenerational effects of prenatal pollution exposure. *LSE Working Paper*.

Colmer, J. M., S. Qin, J. L. Voorheis, and R. Walker (2024). Income, wealth, and environmental inequality in the United States. Working Paper w33050, National Bureau of Economic Research.

Commons, J. (1908). Standardization of housing investigations. *Publications of the American Statistical Association 11*(84), 319–326.

Currie, J., J. Graff Zivin, J. Mullins, and M. Neidell (2014). What do we know about short- and long-term effects of early-life exposure to pollution? *Annual Review of Resource Economics 6*, 217–247.

Currie, J., R. Walker, and J. Voorheis (2023). What caused racial disparities in particulate exposure to fall? New evidence from the clean air act and satellite-based measures of air quality. *American Economic Review 113*(1), 71–97.

Derenoncourt, E., C. H. Kim, M. Kuhn, and M. Schularick (2022). The racial wealth gap, 1860-2020. *Working Paper*.

Derenoncourt, E., C. H. Kim, M. Kuhn, and M. Schularick (2023). Changes in the distribution of black and white wealth since the us civil war. *Journal of Economic Perspectives 37*(4), 71–89.

Deryugina, T., G. Heutel, N. H. Miller, D. Molitor, and J. Reif (2019). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review 109*(12), 4178–4219.

Eisenbud, M. (1978). Levels of exposure to sulfur oxides and particulates in *newyorkcity* and their sources. *Bulletin of the New York Academy of Medicine 54*(11), 991–1011.

Fishback, P., J. LaVoice, A. Shertzer, and R. Walsh (2023). The HOLC maps: How race and poverty influenced real estate professionals' evaluation of lending risk in the 1930s. *Journal of Economic History*.

Fouquet, R. (2011). Long run trends in energy-related external costs. *Ecological Economics 70*(12), 2380–2389.

Graff Zivin, J. and M. Neidell (2013). Environment, health, and human capital. *Journal of Economic Literature 51*(3), 689–730.

Hanlon, W. (2024). London fog: a century of pollution and mortality, 1866–1965. *Review of Economics and Statistics 106*(4), 910–923.

Hanna, R. and P. Oliva (2015). The effect of pollution on labor supply: Evidence from a natural experiment in mexico city. *Journal of Public Economics 122*, 68–79.

Heblich, S., A. Trew, and Y. Zylberberg (2021). East-side story: Historical pollution and persistent neighborhood sorting. *Journal of Political Economy 129*(5), 1508–1552.

Isen, A., M. Rossin-Slater, and W. Walker (2017). Every breath you take—every dollar you'll make: The long-term consequences of the Clean Air Act of 1970. *Journal of Political Economy 125*, 848–902.

Kioumourtzoglou, M., M. Power, J. Hart, O. Okereke, B. Coull, F. Laden, and M. Weisskopf (2017). The association between air pollution and onset of depression among middle-aged and older women. *American Journal of Epidemiology 185*, 801–109.

Lavy, V., A. Ebenstein, and S. Roth (2014). The impact of short term exposure to ambient air pollution on cognitive performance and human capital formation. *National Bureau of Economic Research Working Paper w20648*.

Massey, D. and N. Denton (1993). *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.

Mellon Institute (1914). *Some Engineering Phases of Pittsburgh's Smoke Problem. Smoke Investigation Bulleting No. 8. Mellon Institute of Industrial Research and School of Specific Industries.* Pittsburgh: University of Pittsburgh.

O'Connor, J. J. (1913). *The Economic Cost of the Smoke Nuisance to Pittsburgh. Smoke Investigation Bulletin No. 4. Mellon Institute of Industrial Research and School of Specific Industries*. Pittsburgh: University of Pittsburgh.

Parton, J. (1868). Pittsburg. *Atlantic Monthly 70*(1).

Price, J., K. Buckles, A. Haws, and H. Wilbert (2023, August 10). The census tree, 1870-1880.

Price, J., K. Buckles, J. Van Leeuwen, and I. Riley (2021). Combining family history and machine learning to link historical records: The census tree data set. *Explorations in Economic History 80*, 101391.

Rothstein, R. (2017). *The Color of Law: A Forgotten History of How Our Government Segregated America*. New York: Liveright Publishing.

Rowe, R. D., C. M. Lang, and L. G. Chestnut (1996). Critical factors in computing externalities for electricity resources. *Resource and Energy Economics 18*(4), 363–394.

Ruggles, S., S. Flood, R. Goeken, J. Grover, E. Meyer, J. Pacas, and M. Sobek (2020). IPUMS USA: Version 10.0. https://doi.org/10.18128/D010.V10.0. Minneapolis, MN: IPUMS.

Schlenker, W. and W. Walker (2016). Airports, air pollution, and contemporaneous health. *Review of Economic Studies 83*, 768–809.

Shertzer, A., R. Walsh, and J. Logan (2016). Segregation and neighborhood change in northern cities: New historical GIS data from 1900–1930. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 49*(4), 187–197.

Shertzer, A. and R. P. Walsh (2019). Racial sorting and the emergence of segregation in American cities. *Review of Economics and Statistics 101*(3), 415–427.

Tarr, J. and B. Lamperes (1981). Changing fuel use behavior and energy transitions: The Pittsburgh smoke control movement, 1940–195. *Journal of Social History 14*(4), 561–588.

U.S. Department of Commerce (1917). Abstract of the Census of Manufacturers: 1914. Technical report, Washington, DC.

White, C. W. and C. Marcy (1912). A study of the influences of varying densities of city smoke on the mortality from pneumonia and tuberculosis. *Transactions of the Fifteenth International Congress on Hygiene and Demography 14*(4), 1020–1027.

A Summary Statistics and Additional Results

This appendix presents summary statistics, full sets of coefficients from regressions discussed in the main text, and additional robustness checks. Tables A.1, A.2, A.3 show summary statistics for our primary sample of household heads; men linked between the 1910, 1930, and 1940 Censuses; and Allegheny County employees, respectively.

Table A.5 corresponds with Figure 7 and reports the full set of coefficients of our soot-fall measure regressed on the economic and demographic characteristics of household heads. In addition to the coefficients shown in Figure 7, Table A.5 also includes coefficients for age, sex, and an indicator variable for missing Occupation Income Scores. Age is not significantly correlated with pollution exposure in 1910 and 1930, but it displays a significant and negative correlation with pollution exposure in 1920 and 1940. The coefficient on female begins as positive and statistically significant in 1910 before losing magnitude in 1920 and 1930 and finally becoming significantly negative in 1940. The indicator for missing Occupation Income Score takes on a value of one when an individual reports participating in labor force but has an Occupation Income Score of zero assigned in the IPUMS data (Ruggles et al. (2020). Including this variable partially corrects the measurement error that would otherwise attenuate the relationship between pollution exposure and Occupation Income Score.

As sensitivity analyses, Tables A.6 and A.7 show results from the same regression depicted in Figure 7 and Table A.5 but vary the underlying sample. Rather than household heads, Table A.6 uses single men not living with their parents or other older relatives. Table A.7 uses a sample of women selected by the same criteria. Because married couple predominate in the sample of household heads, one might reasonably wonder if the results shown in Figure 7 are driven by factors specific to married couples. For example, individuals may prioritize spending money on living in cleaner areas after having children compared to when they were single. Running the same regression with other samples allows for a more comprehensive assessment of the relationships between pollution exposure and individual characteristics, and examining men and women separately allows for comparisons between genders.

Table A.6 shows that the patterns of pollution exposure for single men are nearly identical to those of household heads. Greater income, as measured by Occupation Income Score, is associated with lower pollution. Black and foreign-born men are exposed to higher levels of pollution in every year except the anomalous year of 1920, and the black-white gap grows substantially between 1910 and 1940. Speaking English reduces the pollution penalty associated with foreign-born, but the coefficients have lower statistical significance than in the sample of household heads. Table A.7 shows that single women exhibit similar patterns of pollution exposure as well after we drop women working as live-in domestic servants.[19]

Tables A.9 and A.10 expand on results presented Figures 8 and 10. Table A.9 provides evidence that the anomalous patterns of pollution exposure we find around 1920 emerged because of a short-term shift in the sources of air pollution rather than a resorting of residents. Black residents in 1920 were living in areas that had been more polluted in 1910-12 and would be again in 1929-30 and 1938-39, even though those neighborhoods were less polluted in 1923-24. Table A.10 holds pollution constant at its 1910 baseline and allows demographics to shift over time. The increasing magnitude in the black coefficient over time and the corresponding decrease in the foreign-born coefficient suggest that black residents sorted into already heavily polluted neighborhoods between 1910 and 1940 while foreign-born residents exited.

Table A.11 shows the full set of coefficients for the regressions used in Table 8. The sign and significance of coefficients for variables besides race and mover status remain broadly similar to results from A.5. For the linked samples in columns 3 through 6 we are also able to include additional controls for changes in Occupation Income Score and labor force status. Those entering the labor force tend to have higher exposure to pollution in 1940, perhaps indicative of the effect of leaving childhood homes in cleaner areas. Additionally, though the coefficient is not always statistically significant, those whose income increased between the initial Census year and 1940 experienced less pollution in 1940. This relationship suggests that individuals moved into cleaner areas as their incomes increased. Additionally, the coefficients for race and the interaction between race and mover status remaining positive and significant despite controlling for changes in income. The robustness of these coefficients provides further evidence that black residents of Pittsburgh faced barriers to entering cleaner neighborhoods that could not be overcome by increasing income.

[19]Women working as domestic servants varied from the rest of the population because they lived in wealthier areas with their employers despite having very low incomes.

Table A.1: Summary Statistics for Household Head Estimation Sample

	Census 1910	Census 1920	Census 1930	Census 1940
Sootfall	81.44	115.6	84.07	89.81
	(20.68)	(28.04)	(22.00)	(52.55)
Occ. Income Score	19.04	18.38	19.15	20.75
	(15.43)	(15.21)	(14.91)	(14.13)
Black	0.0546	0.0625	0.0832	0.0867
	(0.227)	(0.242)	(0.276)	(0.281)
Foreign Born	0.431	0.373	0.312	0.247
	(0.495)	(0.484)	(0.463)	(0.431)
F.Born, Spk.English	0.356	0.343	0.301	
	(0.479)	(0.475)	(0.459)	
Age	43.03	43.81	45.03	47.33
	(12.92)	(13.12)	(13.22)	(14.14)
Female	0.146	0.148	0.155	0.212
	(0.353)	(0.355)	(0.362)	(0.408)
Ever Married	0.960	0.958	0.947	0.927
	(0.196)	(0.201)	(0.224)	(0.260)
Widowed/Widower	0.138	0.137	0.136	0.163
	(0.345)	(0.343)	(0.343)	(0.370)
Divorced	0.00333	0.00392	0.00805	0.0115
	(0.0576)	(0.0625)	(0.0894)	(0.106)
Owns Home	0.281	0.284	0.403	0.322
	(0.449)	(0.451)	(0.491)	(0.467)
In Labor Force	0.873	0.874	0.861	0.766
	(0.333)	(0.332)	(0.346)	(0.423)
Missing Occ Score	0.194	0.195	0.150	0.0114
	(0.395)	(0.396)	(0.357)	(0.106)
Observations	108460	130055	153327	150042

Table A.2: Summary Statistics for Census Linking

	baseline		linked_1930		linked_1910	
	mean	sd	mean	sd	mean	sd
Pollution	1.710	(1.001)	1.720	(1.000)	1.786	(1.000)
Black	0.085	(0.279)	0.057	(0.233)	0.026	(0.158)
Age	38.153	(16.795)	36.537	(18.213)	51.281	(13.970)
Foreign Born	0.159	(0.366)	0.151	(0.358)	0.155	(0.362)
Never Married	0.351	(0.477)	0.437	(0.496)	0.196	(0.397)
Divorced	0.009	(0.096)	0.006	(0.078)	0.011	(0.105)
Widower	0.084	(0.277)	0.071	(0.257)	0.147	(0.354)
Working	0.517	(0.500)	0.465	(0.499)	0.549	(0.498)
Income	1132.815	(886.470)	1122.851	(1730.091)	1782.624	(1372.912)
Homeowner	0.376	(0.484)	0.451	(0.498)	0.504	(0.500)
Female	0.523	(0.499)	0.503	(0.500)	0.463	(0.499)
Observations	415505		297565		85363	

Note: Table reports 1940 summary statistics of males from the 1940 census. The "F.Born,Spk.English" variable is omitted because data was not collected on English fluency for the 1940 census.

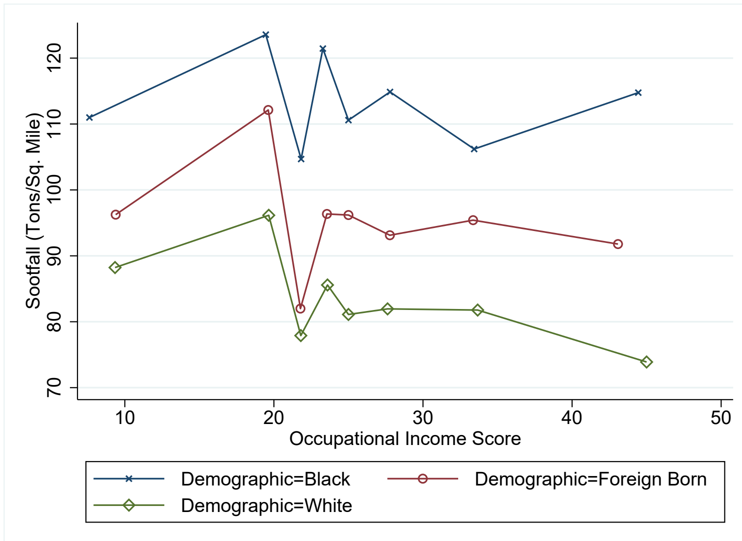
Table A.3: Summary Statistics for Matched County Employees

	Observations	Mean	SD	Min	Max
Panel A: 1910					
Sootfall	224	79.305	21.748	49.296	126.884
Annual Wage Income	224	1354.014	1024.251	300.000	9999.960
Occupation Income Score	224	25.884	9.813	6.000	62.000
Black	224	0.036	0.186	0.000	1.000
Foreign Born	224	0.121	0.326	0.000	1.000
Panel B: 1920					
Sootfall	235	112.131	27.069	75.900	202.066
Annual Wage Income	235	1706.752	789.465	660.000	7999.920
Occupation Income Score	235	28.251	8.980	15.000	80.000
Black	235	0.089	0.286	0.000	1.000
Foreign Born	235	0.102	0.303	0.000	1.000

Table A.4: Matching Statistics for Hedonic Data

	1930	1940
Census Data		
Single Head Households	154,803	173,035
Useable Addresses	74,400	75,962
Assessor's Data		
Single Family (pre 1930.1940)	66,260	83,617
Usable Addresses	65,566	82,688
Merged Data		
Strict Address Matches	19,844	22,664
Assessor's Lat/Long in E.D./Tract	19,013	20,453
Clean Hedonic Variables	16,408	17,641

Figure A.1: Pollution vs. Income by Demographic Group (1940 using Occupational Income Scores)



Note: This figure reports binned means of sootfall exposure, by demographics (White, Foreign Born, Black) for 1940 where the bins are constructed based on Occupational Income Score. Bin boundaries were chosen to equate populations across bins.

Figure A.2: Annual Steel Production in Pennsylvania

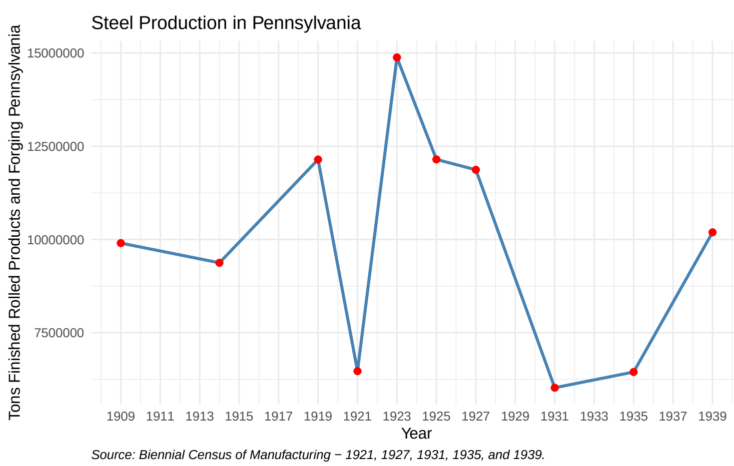


Table A.5: Pollution Exposure for Household Heads

	Census 1910	Census 1920	Census 1930	Census 1940
Occscore	-0.0508***	-0.0352*	-0.129***	-0.0886***
	(0.0131)	(0.0196)	(0.0203)	(0.0180)
Black	0.278***	-0.282***	0.352***	0.636***
	(0.101)	(0.0646)	(0.0718)	(0.242)
Foreign Born	0.539***	0.144	0.526***	0.401***
	(0.0946)	(0.109)	(0.117)	(0.0680)
Speaks English	-0.254***	-0.155*	-0.302***	
	(0.0893)	(0.0909)	(0.0942)	
Home Owned	-0.306***	0.0188	-0.207***	-0.226***
	(0.0384)	(0.0388)	(0.0345)	(0.0502)
Worker	0.152***	0.119***	0.282***	0.0701**
	(0.0326)	(0.0446)	(0.0505)	(0.0350)
Widower	0.0362*	0.0207	0.0537***	0.0451***
	(0.0211)	(0.0210)	(0.0172)	(0.0156)
Divorced	0.125*	-0.0551	-0.0126	-0.0295
	(0.0668)	(0.0508)	(0.0343)	(0.0392)
Female	0.103***	-0.0446**	0.00590	-0.101***
	(0.0270)	(0.0191)	(0.0194)	(0.0256)
Age	0.000486	-0.00195***	0.000834	-0.00314***
	(0.000686)	(0.000745)	(0.000644)	(0.000812)
Missing Occ.	-0.191***	-0.0934*	-0.341***	-0.0991
	(0.0404)	(0.0527)	(0.0544)	(0.0675)
Ever Married	-0.171***	0.138***	-0.0876***	-0.118*
	(0.0353)	(0.0294)	(0.0255)	(0.0622)
Constant	3.902***	4.051***	3.787***	1.998***
	(0.0683)	(0.0579)	(0.0610)	(0.128)
Observations	108460	130055	153327	150042
R^2	0.063	0.008	0.045	0.077
Adjusted R^2	0.063	0.007	0.045	0.077

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Pollution levels and Occupation Scores are expressed in terms of within census standard deviations. SEs clustered at the Census division.

Table A.6: Pollution Exposure for Single Men

	Census 1910	Census 1920	Census 1930	Census 1940
Occscore	-0.0443***	-0.0352**	-0.105***	-0.0626***
	(0.0134)	(0.0139)	(0.0138)	(0.0138)
Age	0.00626**	0.000326	0.00285***	0.000446
	(0.00293)	(0.000951)	(0.000820)	(0.00107)
Black	0.266**	-0.168**	0.362***	0.773**
	(0.133)	(0.0710)	(0.0857)	(0.326)
Foreign Born	0.433**	0.374*	0.382***	0.426***
	(0.189)	(0.191)	(0.128)	(0.0851)
Speaks English	-0.242	-0.214	-0.197*	
	(0.207)	(0.171)	(0.108)	
Home Owned	-0.376***	-0.0951**	-0.250***	-0.269***
	(0.0523)	(0.0414)	(0.0385)	(0.0722)
Widower	-0.189**	0.0143	-0.0941***	-0.150***
	(0.0734)	(0.0305)	(0.0296)	(0.0533)
Divorced	-0.0187	-0.0157	-0.102***	-0.139**
	(0.0898)	(0.0695)	(0.0387)	(0.0607)
Worker	0.213***	0.235***	0.336***	0.133***
	(0.0780)	(0.0604)	(0.0565)	(0.0370)
Constant	3.513***	4.130***	3.960***	1.676***
	(0.120)	(0.0781)	(0.0669)	(0.0947)
Observations	30904	25330	27060	24298
R^2	0.076	0.021	0.059	0.128
Adjusted R^2	0.075	0.021	0.059	0.128

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Pollution levels and Occupation Scores are expressed in terms of within census standard deviations. SEs clustered at the Census division.

Table A.7: Pollution Exposure for Single Women

	Census 1910	Census 1920	Census 1930	Census 1940
Occscore	-0.0474***	0.0400**	-0.0704***	-0.0628***
	(0.0151)	(0.0187)	(0.0175)	(0.0178)
Age	-0.00256	-0.000379	0.000472	-0.00460***
	(0.00164)	(0.00101)	(0.000857)	(0.000982)
Black	0.276***	-0.0911	0.547***	0.805***
	(0.0985)	(0.0811)	(0.105)	(0.296)
Foreign Born	0.554***	0.216**	0.395***	0.391***
	(0.0962)	(0.0973)	(0.0975)	(0.0634)
Speaks English	-0.375***	-0.151*	-0.202**	
	(0.0919)	(0.0810)	(0.0867)	
Home Owned	-0.231***	0.0187	-0.182***	-0.167***
	(0.0356)	(0.0387)	(0.0356)	(0.0535)
Widow	0.0299	0.0662*	0.0715**	0.0618**
	(0.0451)	(0.0348)	(0.0282)	(0.0272)
Divorced	0.134**	0.0403	0.0111	-0.0347
	(0.0605)	(0.0620)	(0.0402)	(0.0411)
Worker	0.138**	-0.106**	0.187***	0.0643
	(0.0535)	(0.0467)	(0.0501)	(0.0430)
Constant	4.022***	4.036***	4.096***	1.912***
	(0.144)	(0.0625)	(0.0695)	(0.109)
Observations	21401	23964	30638	35049
R^2	0.048	0.004	0.036	0.086
Adjusted R^2	0.048	0.004	0.035	0.086

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Pollution levels and Occupation Scores are expressed in terms of within census standard deviations. Live-in domestic servants are dropped from the sample. SEs clustered at the Census division.

Table A.8: Pollution Exposure HH Heads w/ Southern Migrants

	Census 1910	Census 1920	Census 1930	Census 1940
Occscore	-0.0731***	-0.0422*	-0.117***	-0.0795***
	(0.0149)	(0.0215)	(0.0169)	(0.0163)
Southern Born	-0.0399	-0.211***	-0.117***	0.120*
	(0.0405)	(0.0568)	(0.0308)	(0.0657)
Black	0.430***	-0.225***	0.348***	0.602**
	(0.105)	(0.0684)	(0.0742)	(0.274)
Southern Born X Black	-0.201***	0.117*	0.130***	-0.0602
	(0.0623)	(0.0656)	(0.0421)	(0.0946)
Home Owned	-0.272***	0.0324	-0.208***	-0.224***
	(0.0399)	(0.0442)	(0.0377)	(0.0494)
Worker	0.211***	0.131***	0.226***	0.0455
	(0.0371)	(0.0494)	(0.0410)	(0.0345)
Widower	0.0524**	-0.00178	0.0747***	0.0513***
	(0.0253)	(0.0250)	(0.0190)	(0.0162)
Divorced	0.0619	-0.0942*	-0.00970	-0.0306
	(0.0751)	(0.0511)	(0.0352)	(0.0359)
Female	0.124***	-0.0613***	-0.0138	-0.0942***
	(0.0324)	(0.0218)	(0.0218)	(0.0239)
Age	0.00184**	-0.00177**	0.000880	-0.00355***
	(0.000809)	(0.000865)	(0.000733)	(0.000850)
Missing Occ.	-0.253***	-0.120**	-0.316***	-0.0803
	(0.0439)	(0.0567)	(0.0458)	(0.0555)
Ever Married	-0.173***	0.152***	-0.0999***	-0.107**
	(0.0306)	(0.0318)	(0.0281)	(0.0500)
Constant	3.827***	4.044***	3.826***	2.002***
	(0.0661)	(0.0610)	(0.0663)	(0.126)
Observations	61679	81481	105552	112939
R^2	0.042	0.011	0.047	0.085
Adjusted R^2	0.042	0.011	0.047	0.085

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Pollution levels and Occupation Scores are expressed in terms of within census standard deviations. Foreign born individuals are dropped. SEs clustered at the Census division.

Table A.9: Fixing 1920 Demographics

	Pollution 1910-1912	Pollution 1923-1924	Pollution 1929-1930	Pollution 1938-39
Occscore	-0.0945***	-0.0352*	-0.0976***	-0.0630***
	(0.0170)	(0.0196)	(0.0177)	(0.0155)
Black	0.332***	-0.282***	0.214***	0.390***
	(0.0912)	(0.0646)	(0.0709)	(0.115)
Foreign Born	0.368***	0.144	0.407***	0.446***
	(0.0987)	(0.109)	(0.115)	(0.0405)
Speaks English	-0.0239	-0.155*	-0.120	
	(0.0842)	(0.0909)	(0.0962)	
Home Owned	-0.298***	0.0188	-0.216***	-0.297***
	(0.0387)	(0.0388)	(0.0378)	(0.0361)
Worker	0.217***	0.119***	0.204***	0.128***
	(0.0413)	(0.0446)	(0.0426)	(0.0378)
Widower	0.0519**	0.0207	0.0414**	0.0481***
	(0.0201)	(0.0210)	(0.0199)	(0.0168)
Divorced	0.0990*	-0.0551	0.0449	-0.0307
	(0.0521)	(0.0508)	(0.0449)	(0.0503)
Female	0.0248	-0.0446**	-0.00756	-0.0436**
	(0.0190)	(0.0191)	(0.0184)	(0.0202)
Age	-0.000266	-0.00195***	-0.000684	-0.00342***
	(0.000644)	(0.000745)	(0.000659)	(0.000521)
Missing Occ.	-0.249***	-0.0934*	-0.278***	-0.178***
	(0.0501)	(0.0527)	(0.0516)	(0.0472)
Ever Married	-0.118***	0.138***	-0.0709***	-0.0863***
	(0.0293)	(0.0294)	(0.0255)	(0.0289)
Constant	3.850***	4.051***	3.960***	1.966***
	(0.0580)	(0.0579)	(0.0581)	(0.0678)
Observations	129287	130055	130055	130055
R^2	0.058	0.008	0.041	0.076
Adjusted R^2	0.058	0.007	0.041	0.076

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Pollution levels and Occupation Scores are expressed in terms of within census standard deviations. SEs clustered at the Census division.

Table A.10: Fixing 1910 Pollution

	Census 1910	Census 1920	Census 1930	Census 1940
Occscore	-0.0508***	-0.0945***	-0.138***	-0.140***
	(0.0131)	(0.0170)	(0.0215)	(0.0241)
Black	0.278***	0.332***	0.346***	0.434**
	(0.101)	(0.0912)	(0.0951)	(0.182)
Foreign Born	0.539***	0.368***	0.427***	0.133**
	(0.0946)	(0.0987)	(0.0946)	(0.0660)
Speaks English	-0.254***	-0.0239	-0.213***	
	(0.0893)	(0.0842)	(0.0758)	
Home Owned	-0.306***	-0.298***	-0.196***	-0.152**
	(0.0384)	(0.0387)	(0.0374)	(0.0592)
Worker	0.152***	0.217***	0.288***	0.190***
	(0.0326)	(0.0413)	(0.0536)	(0.0518)
Widower	0.0362*	0.0519**	0.0800***	0.0523***
	(0.0211)	(0.0201)	(0.0191)	(0.0160)
Divorced	0.125*	0.0990*	0.0271	0.0911*
	(0.0668)	(0.0521)	(0.0457)	(0.0511)
Female	0.103***	0.0248	-0.0300	-0.0956***
	(0.0270)	(0.0190)	(0.0214)	(0.0245)
Age	0.000487	-0.000266	-0.000419	-0.000791
	(0.000686)	(0.000644)	(0.000643)	(0.000850)
Missing Occ.	-0.191***	-0.249***	-0.372***	-0.307***
	(0.0404)	(0.0501)	(0.0554)	(0.0706)
Ever Married	-0.171***	-0.118***	-0.107***	-0.161***
	(0.0353)	(0.0293)	(0.0278)	(0.0436)
Constant	3.902***	3.850***	3.860***	4.099***
	(0.0683)	(0.0580)	(0.0627)	(0.101)
Observations	108460	129287	139778	139547
R^2	0.063	0.058	0.044	0.045
Adjusted R^2	0.063	0.058	0.044	0.045

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Pollution levels and Occupation Scores are expressed in terms of within census standard deviations. SEs clustered at the Census division.

Table A.11: Link Between Moving and Pollution Exposure (Full Results)

	1940 Full	1940 Full	1940 - 1930	1940 - 1910	1940 - 1930	1940 - 1910
Black	0.553**	0.271*	0.190	0.0268	-1.285	-5.084
	(0.217)	(0.141)	(0.130)	(0.157)	(3.685)	(4.313)
Age	-0.00251***	-0.00273***	-0.00429***	-0.00445***	-0.0784***	-0.00884
	(0.000919)	(0.000942)	(0.00108)	(0.00105)	(0.0292)	(0.0351)
Foreign Born	0.366***	0.346***	0.366***	0.350***	6.120***	4.402**
	(0.0623)	(0.0568)	(0.0600)	(0.0593)	(1.683)	(1.716)
Never Married	0.0607***	0.0452*	0.0227	0.0637***	0.173	1.007
	(0.0229)	(0.0233)	(0.0186)	(0.0229)	(0.596)	(0.780)
Divorced	-0.0407	-0.0340	-0.00926	-0.0441	-1.065	0.970
	(0.0375)	(0.0352)	(0.0372)	(0.0375)	(1.450)	(1.439)
Widower	0.0280**	0.0281***	0.0314**	0.0354**	-0.0527	0.175
	(0.0108)	(0.0104)	(0.0124)	(0.0156)	(0.380)	(0.560)
Working	0.0683***	0.0683***	0.112***	0.0482**	-0.0880	-1.350
	(0.0144)	(0.0138)	(0.0238)	(0.0244)	(0.728)	(1.061)
Income 1940	-0.0950***	-0.0909***				
	(0.0157)	(0.0150)				
Homeowner	-0.207***	-0.230***	-0.248***	-0.211***	-3.975**	-4.972***
	(0.0463)	(0.0420)	(0.0457)	(0.0436)	(1.586)	(1.569)
Female	-0.0769***	-0.0730***	-0.0422***	-0.113***	-0.276	-1.118
	(0.0149)	(0.0140)	(0.00985)	(0.0239)	(0.324)	(0.745)
Change Nbrhd		-0.100***	-0.202***	-0.362***	-8.856***	-17.43***
		(0.0284)	(0.0382)	(0.0653)	(2.474)	(2.547)
Black X Chg Nbrhd		0.418***	0.314***	0.455***	10.23***	18.10***
		(0.130)	(0.0909)	(0.120)	(3.286)	(4.010)
FB X Chg Nbrhd		0.0100	-0.0543	-0.0646	-4.469***	-7.963***
		(0.0275)	(0.0438)	(0.0556)	(1.595)	(2.145)
Change in OccInc			-0.00239***	-0.00415***	-0.0216	-0.178***
			(0.000842)	(0.00114)	(0.0292)	(0.0440)
Enters Workforce			0.112***	0.0315	0.275	4.215***
			(0.0253)	(0.0281)	(0.752)	(1.133)
Exits Workforce			0.163***	0.0851**	-2.353	-0.0585
			(0.0405)	(0.0426)	(1.600)	(1.445)
Entry X Occ2nd			-0.00781***	-0.00645***	0.0245	-0.183***
			(0.00189)	(0.00173)	(0.0525)	(0.0644)
Exit X Occ1st			-0.00436***	-0.00407**	0.0744	-0.0460
			(0.00153)	(0.00168)	(0.0620)	(0.0570)
OCC1st for ChgOcc			-0.00505***	-0.00471***	-0.00485	-0.0214
			(0.00107)	(0.00117)	(0.0346)	(0.0438)
Missing OCC Both			0.0396	0.0729	2.629	4.198
			(0.0821)	(0.152)	(2.812)	(6.592)
Occ2nd for Missing Occ1st			-0.00242***	0.000473	-0.0388*	0.0492*
			(0.000688)	(0.000644)	(0.0223)	(0.0270)
Occ1st for Missing Occ2nd			-0.00107	0.00369	0.162**	0.234***
			(0.00163)	(0.00264)	(0.0754)	(0.0884)
Constant	1.815***	1.822***	1.978***	2.369***	9.398**	12.33***
	(0.106)	(0.113)	(0.125)	(0.158)	(4.078)	(4.535)
Dep. Var.	1940 Soot	1940 Soot	1940 Soot	1940 Soot	Chg Soot Pctile	Chg Soot Pctile
N	415505	415505	297565	85363	297565	85363

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Pollution levels and incomes are expressed in terms of within census standard deviations. Columns 1 and 2 include all individuals in the 1940 census, with changes in neighborhood and/or county being based on the census' 5 year move move variable. Columns 3 and 4 match samples of men from the 1940 census to the 1930 and 1910 censuses respectively. Columns 5 and 6 use the matched samples as well, with the dependent variable being the change in pollution percentile from period 1 to period 2. SEs clustered at the Census division.