East-Side Story: Historical Pollution and Persistent Neighborhood Sorting

Stephan Heblich

University of Toronto

Alex Trew

University of Glasgow

Yanos Zylberberg

University of Bristol

Why are the east sides of formerly industrial cities more deprived? To answer this question, we use individual-level census data and create historical pollution patterns derived from the locations of 5,000 industrial chimneys and an atmospheric model. We show that this observation results from path-dependent neighborhood sorting that began during the Industrial Revolution, as prevailing winds blew pollution eastward. Past pollution explains up to 20% of observed neighborhood segregation in 2011, even though coal pollution stopped in the 1970s. We develop a quantitative model to identify the role of neighborhood effects and relocation rigidities underlying this persistence.

I. Introduction

Cities that were formerly reliant on industry tend today to have eastern suburbs that are notably poorer than their western suburbs. This

observation is echoed in media stories about the east side of London, New York, or Paris and in popular culture (such as in the long-running BBC soap opera *EastEnders*). We show that the east-west gradient is partially a remnant of the distribution of the atmospheric pollution that affected cities during the Industrial Revolution. Pollution from historical factories accounted for about 15% of the variation in neighborhood composition in 1881. There is no evidence of excess deprivation in neighborhoods downwind from industrial chimneys before the rise of industrial coal in 1817. Industrial coal pollution effectively stopped in the 1970s, but the path dependence in neighborhood sorting is still felt today. To understand this, we develop a quantitative model of neighborhood sorting and provide new evidence on how the combination of neighborhood effects and relocation frictions can generate tipping-like dynamics.

This paper is the first to present a long-run analysis of the effects of pollution on the internal structure of cities. While the impact of pollution on welfare in cities is often highlighted in modern policy debates, the long-run consequences for economic agents are less well known. Providing such evidence is challenging since systematic air-pollution monitoring on a fine spatial scale started only after industrial coal pollution markedly slowed down. To fill this gap, we develop a novel method of modeling historical pollution within English cities at the end of the nineteenth century. Specifically, we geolocate industrial chimneys from historical Ordnance Survey (OS) maps of the 70 largest metropolitan areas in England over the period 1880–1900 and use an atmospheric dispersion modeling system (ADSM) to construct pollution maps from the chimney locations.[1]

[1] To account for heterogeneity in local conditions, the model incorporates additional information on terrain, prevailing winds, and chimney characteristics (chimney dimensions, exit velocity, coal burning temperature).

We combine this measure of late-nineteenth-century pollution exposure with unique, neighborhood-level panel data spanning nearly 200 years. The latter is made possible by a newly developed algorithm that helps us geolocate individual addresses from the 1881 census and assign them to 5,500 low-level administrative units.

We find a strong effect of air pollution on the share of low-skilled workers in 1881. A pollution differential equivalent to that between the 10th and 90th percentiles in Manchester would be associated with a gradient of 16 percentage points in the share of low-skilled workers. Importantly, there is no excess deprivation in neighborhoods downwind from industrial chimneys prior to the rise of industrial coal use. Moreover, we continue to find substantial persistence in the effect of historical pollution on within-city distribution of low-skilled workers, even after the second Clean Air Act of 1968 abruptly decreased pollution from coal burning.[2] The previous 10th–90th-percentile difference would explain a similar gradient in neighborhood composition in 2011 and a 40% difference in property prices. We also find that the dynamics of persistence between 1971 and 2011 show evidence of nonlinearities, with mean reversion for intermediate values of within-city pollution and inertia for neighborhoods with values of within-city pollution at the high and low tails of pollution exposure.

An immediate identification concern is that industrial chimneys are not randomly allocated within cities. We address this issue in two ways. First, we condition our analysis on the distance to industrial chimneys and analyze neighborhood composition in all different directions relative to the chimney. This spatial differencing exercise reveals excess deprivation in 1881 along a narrow corridor downwind of industrial chimneys. Second, we present an instrumental variable (IV) strategy to address the concern that chimneys may have been selectively located upwind of poor areas. Specifically, we instrument the pollution pattern induced by actual chimneys with a predicted pollution pattern that exploits exogenous locations of pollution sources. The choice of exogenous pollution sources rests on the fact that steam engines need water for cooling (Maw, Wyke, and Kidd 2012). We exploit waterways in 1827 as an exogenous location factor that predicts the actual pollution pattern. As we also condition on distance to the waterways and exclude neighborhoods bordering the waterside, we exploit the difference between upwind and downwind neighborhoods at the same distance from potential factories located along waterways. The IV specification delivers qualitatively similar results as the baseline ordinary least squares (OLS) specification.

---

[2] The first Clean Air Act was enacted in 1956 as a reaction to the Great Smog of 1952 in London. However, the second Clean Air Act in 1968 caused a much more pronounced drop in coal consumption.

To interpret the persistence of past pollution, we develop a dynamic model of neighborhood choice with two types of households—low-skilled and high-skilled—that differ only in their income. Our main contribution relative to previous studies lies in the way we model moving rigidities. We suppose that, in each period, households are subject to an exogenous relocation shock. Conditional on being able to move, the decision to actually do so reflects households' relative preferences for neighborhoods (as in, e.g., Bayer et al. 2016). This setup yields a simple dynamic equation that characterizes the relative demand for neighborhoods. A key advantage is the comparatively low data requirement; instead of individual data or flows between neighborhoods, identification requires only the aggregate share of low-skilled households in each neighborhood over 3 periods. Using the shares of low-skilled households from 1971 to 2011, we estimate the relative demand for neighborhoods using the generalized method of moments. To address endogeneity concerns, we show how to use historical pollution as an instrument for current neighborhood composition and its subsequent evolution. We find that the persistence of neighborhood sorting is tied not only to relocation frictions but also to its interaction with preferences for neighborhood composition. Since past sorting is partially inherited, a backward-looking element of persistence is captured in the estimate of the share of residents who move in each period. An additional effect arises from the forward-looking behavior of movers: they anticipate the slow future adjustment of neighborhood composition in their current valuation of a neighborhood.

We use our model to undertake counterfactuals. In particular, we can quantify the fraction of present-day sorting due to pollution and social housing policies. In the most-polluted cities, historical pollution increased neighborhood segregation by about 1 standard deviation (SD), as evaluated in 1971 as well as 2011. We also show that the liberalization of social housing contributed to the persistence of neighborhood sorting: fixing social housing at its level and distribution in 1971 would have reduced segregation markedly by preventing well-connected neighborhoods from further gentrifying.

Our paper contributes to three different strands of literature. First, our work is related to that of Lee and Lin (2018), who look at exogenous natural amenities as a driver of neighborhood sorting. We, however, study the consequences of a temporary disamenity. To the best of our knowledge, we are the first to show that the high level of temporary pollution from industrial coal use modified the spatial organization of cities in the long run.[3] Related studies that look at pollution-induced sorting in a purely

---

[3] Two recent papers find similar patterns of persistence for (i) historical marshes in New York City (Villarreal 2014) and (ii) historical streetcar lines in Los Angeles County (Brooks and Lutz 2019). A large body of literature discusses the path dependence in economic activity across cities (e.g., Davis and Weinstein 2002; Bleakley and Lin 2012).

modern setting include those by Chay and Greenstone (2005) and Banzhaf and Walsh (2008).[4] Kuminoff, Smith, and Timmins (2013) provide a broader review of the residential sorting literature. Our argument further relates to that of Depro, Timmins, and O'Neil (2015), who argue that neighborhood sorting, rather than environmental injustice, is the reason why poor households are more exposed to environmental disamenities. Finally, our work shares a common theme with that of Hanlon (2019), who argues that coal-based pollution was a significant disamenity, with a strong negative impact on city size in England during industrialization, and Chen, Oliva, and Zhang (2017) and Freeman et al. (2019), who document a similar correlation between pollution and residential choices between cities in China.

Second, we relate to a literature on the gentrification of historic centers in US cities (Brueckner and Rosenthal 2009; Guerrieri, Hartley, and Hurst 2013; Baum-Snow and Hartley 2020; Couture and Handbury 2020), which, in turn, builds on previous research on the dynamics of segregation and tipping points (Schelling 1971; Card, Mas, and Rothstein 2008; Logan and Parman 2017). In our context, we mostly identify a social component behind segregation (in contrast to the literature on the United States, which mostly focuses on ethnic considerations). We also present a novel strategy to estimate a dynamic model of neighborhood choice with low data requirements (in contrast to Bayer et al. 2016). Our estimates point to nonlinearities in the dynamics of segregation: highly polluted neighborhoods repel high-income residents even after pollution has waned.

Third, we contribute to quantitative research in economic history. We introduce an algorithm that geolocates census entries in 1881. This could be applied to any historical census in most developed countries. The algorithm exploits the clustering among census entries to infer the geolocation of all residents from a fraction of well-matched neighbors. Another contribution is digitizing historical maps to create a new data set on pollution and the structure of cities. Related to this approach is work by Siodla (2015) and Hornbeck and Keniston (2017), who use historical maps to understand the effects of the great fires in Boston and San Francisco, and Redding and Sturm (2016), who use maps to document destruction in London during the Second World War.

The remainder of the paper is organized as follows. Section II provides elements of context and describes data sources. The reduced-form

---

[4] There is also a broad body of literature on pollution exposure and its effect on productivity (Graff Zivin and Neidell 2012), cognitive performance (Lavy, Ebenstein, and Roth 2016), violent crime (Herrnstadt et al. 2019), and health (Graff Zivin and Neidell 2013; Deryugina et al. 2019; Anderson 2020) that relates to our research. Closely related historical assessments of the effect of coal use on health include Barreca, Clay, and Tarr (2014), Clay, Lewis, and Severnini (2016), and Beach and Hanlon (2018).

evidence on neighborhood sorting and its persistence is discussed in section III. Section IV develops a dynamic model of residential choice. The identification and estimation of the demand for neighborhoods are discussed in section V. Finally, section VI concludes.

II. Historical Background and Data

This section describes the historical setting, our main data sources, the construction and validation of atmospheric pollution between 1880 and 1900, and neighborhood composition in 1817, 1881, and 1971–2011. We also provide suggestive evidence about the role of wind direction in generating spatial inequalities within cities.

A. Historical Background

The start of the classical Industrial Revolution dates to around 1760, marked by the arrival of new technologies in key growth sectors such as textiles, iron, and steam. However, important consequences of that revolution were not realized until much later, and per capita growth rates did not accelerate until after 1830 (Crafts and Harley 1992). Economic growth was accompanied by an energy transition, with coal emerging as the dominant energy source around 1840.[5] Figure 1A shows this rapid energy transition and subsequent trends in coal use. There is a sharp acceleration of coal consumption between 1850 and 1910, which flattens out as electricity and oil rise in importance. The 1926 strike in support of the coal miners reflects the declining importance of coal, even though coal consumption remains relatively stable in absolute terms until the mid-1950s. After the 1952 London Smog caused approximately 4,000 deaths, a first Clean Air Act was passed in 1956, which marked a turning point in air-quality regulation worldwide (Brimblecombe 2006). Once in place, there was an increasing societal understanding of the importance of air quality, which led to the more restrictive 1968 Clean Air Act. The Clean Air Acts penalized the emissions of grit, dust, and "dark smoke" in cities and placed minimum height restrictions on chimneys. Industry subsequently shifted away from coal to the use of cleaner energy sources such as oil, gas, and electricity generated by power stations outside of cities. As apparent in figure 1, these regulations, and particularly the Clean Air Act of 1968, had an immediate and marked impact on coal consumption.[6]

---

[5] As Musson (1976) shows, power derived from waterwheels remained important to industry in the early nineteenth century—steam power was not prevalent outside of textiles until after the 1870s.

[6] The early twentieth century saw a consolidation of industry, with employment peaking at 46% in 1950 (Crafts 2014). The decline in coal consumption preceded the massive deindustrialization, which occurred most rapidly in the 1980s, when state-owned industries were privatized.
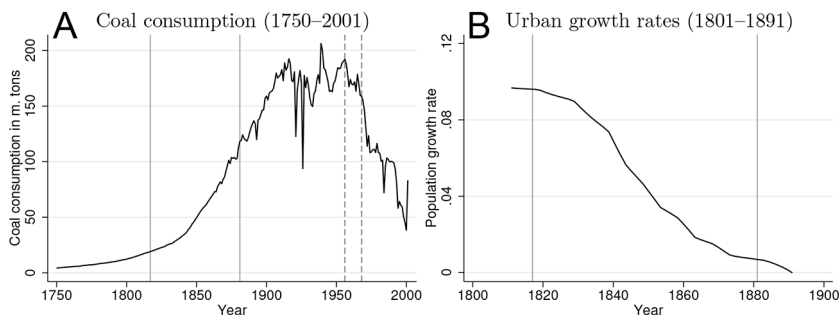
FIG. 1.—Coal consumption and migration during the Industrial Revolution. *A*, Increase and decrease in coal consumption over the period 1750–2000, based on the work of Warde (2007), who reports coal consumption in petajoule. To convert numbers from petajoule to tons, we use a conversion factor of 1 : 34,140. The solid vertical lines indicate the years 1817 and 1881, while the dashed vertical lines mark the introduction of the 1956 and 1968 Clean Air Acts. *B*, Average decadal population growth rate for the period 1801–91 in cities in our sample.

The heavy reliance on coal between 1850 and 1960 generated unprecedented concentrations of sulfur dioxide (SO_2) in the atmosphere, which scarred cities and their surroundings.[7] Mosley (2013), for instance, conjectured a relationship between historical pollution and neighborhood sorting: "In Manchester, prevailing and strongest winds [blow] from the south west. This meant that when the dense sulphurous smoke left Manchester's tall chimneys it usually moved north east, and this was to have a marked effect on the shaping of the city. . . . The poorest city dwellers were forced to live amongst the mills and factories in north-easterly districts . . . the better-paid among Manchester's working classes might at least escape the worst of the smoke" (30–31). The negative impact of atmospheric pollution is also captured in a well-known case of micro-evolutionary change. The dominant form of the peppered moth (*Biston betularia*) at the start of the nineteenth century was the lighter form (*insularia*), as it was camouflaged against predation when on light trees and lichens. The first sightings of the darker form of the moth (*carbonaria*) in the industrial north of England were not until after 1848 (Cook 2003). As pollution caused trees to blacken under layers of soot, the *carbonaria* emerged as the dominant form by the end of the nineteenth century. The decline in air pollution after the Clean Air Acts led to a rapid recovery of the *Biston betularia insularia* after 1970 (Cook 2003).

Along with the structural transformation of the economy, the end of the eighteenth century saw rapid urbanization, with workers from the

[7] We do not attempt to discuss the growing body of literature on pollution and individual health. Instead, we refer to a survey article by Graff Zivin and Neidell (2013) and, specifically, to historical assessments of the effect of coal use on health discussed by Barreca, Clay, and Tarr (2014), Clay, Lewis, and Severnini (2016), and Beach and Hanlon (2018).

countryside flocking into the emerging industrial cities (Shaw-Taylor and Wrigley 2014). As shown in figure 1*B*, the growth of cities started to decline after 1830 and steadily slowed down as the nineteenth century proceeded. By the end of the nineteenth century, the large cross-country migratory flows that marked the early Industrial Revolution had moderated significantly.[8]

This great movement into cities came with an increase in density that overwhelmed Victorian cities. The characteristic back-to-back houses were put together with thin walls and no foundation or ventilation and located within walking distance to the new factories. The lack of suitable housing along with a limited supply of clean water and sanitation created unhealthy urban slums plagued by diseases such as cholera and typhoid, leading to notoriously low life expectancy (Clark 1962). Only in the second half of the nineteenth century did the Public Health Acts of 1872 and 1875 begin to improve the living conditions of the poor.

While poor working-class families were typically stuck in the inner residential areas (or urban slums) around the city center, middle- and, subsequently, lower-middle-classes families started separating their place of work and place of residence, thus encouraging new housing development in suburbs at the fringe of the city. Wealthier suburbs were characterized by private residential gardens and spacious villas, while poorer suburbs were made up of long terraces of byelaw housing. As discussed by Heblich, Redding, and Sturm (2020), the rise of the railways and the subway facilitated suburbanization, with residential development spreading widely around London. However, this wide-sprawling development was specific to London. In other British cities, suburbs remained in walking distance from workplaces (Kellett 1969; Lawton 1972).

The empirical analysis relies on the following observable characteristics of neighborhoods: (i) neighborhood composition in 1817, before the acceleration in coal consumption and around the decline in rural migration to urban centers; (ii) atmospheric pollution, neighborhood composition, and urban structure around 1880–1900, slightly before the peak in coal consumption; and (iii) neighborhood characteristics between 1971 and 2011, after the abrupt decrease in atmospheric pollution.

B. *Data Sources and Construction*

This section provides a summary of the different data sources; a comprehensive description can be found in appendix C (apps. A–H are available online).

---

[8] Ravenstein (1885) and Williamson (1990) show that the portion of city growth due to migration declines over the nineteenth century; by 1881, 75% of individuals in England and Wales resided in the county of their birth.

1.   OS Maps and Geolocation of Pollution Sources

Data on city structure and pollution sources are drawn from the OS maps, which are scaled at 25 inches to 1 mile. These maps are the most detailed topographic maps to cover England and Wales at the turn of the nineteenth century. We restrict the analysis to the wave of maps published between 1880 and 1900 and to the 70 largest metropolitan areas in England at the time (as derived from the 1907 Census of Production). These cities constitute a quasi-exhaustive snapshot of industry and cover 60% of the total population in 1801 and 66% in 2011. The maps contain details on roads, railway, rivers, canals, and public amenities as well as the outline of each building and their use.[9] Most useful for our purposes, these maps mark the locations of factory chimneys, in a sign of the fastidiousness of Victorian mappers. We georeference more than 5,000 chimneys, all matched to a description of the associated workshop or factory. Figure 2 gives some examples of the variety of symbols that were used to mark a chimney on a map; figure 3 depicts our method of extracting information.

   In order to account for sectoral differences in coal use, we extract industrial information along with chimney locations.[10] Aggregate measures of coal use per worker are calculated for the following 11 industrial categories: brick factories, foundries, chemical factories, mining, breweries, tanneries, food processing, textile production, paper production, shipbuilding, and wood processing (Hanlon 2019). A simple textual analysis based on a few keywords to associate a category to the map's description of each industrial site allows us to match 90% of the 5,000 chimneys. The remaining 10% are classified under a generic category ("other manufactures"). The estimated pollution emission E_i from a chimney in industry i is constructed as

$$E_i = \frac{C_i \times L_i}{\text{Ch}_i},$$

where C_i is the industry-specific measure of coal use per worker, L_i is total employment in industry i, and Ch_i is the total number of chimneys of type i. Accordingly, L_i/Ch_i gives us the average number of workers per chimney in industry i.[11]

---

[9] We use the OS maps to extract information about the city contour—the location of town halls, market halls, churches, schools, universities, parks, theaters, museums, churches, and hospitals.

[10] The industrial category—measured with error—is the only information that we use in order to proxy for the average pollution emission associated with a chimney. The measurement error induced by our (imperfect) modeling of pollution emissions is likely to generate an attenuation bias. We show, however, in a robustness check that the industry weight already contains significant information.

[11] We report the estimated pollution emission per industry, E_i, in table A1 (tables A1–A13 are available online). As chimneys in the unidentified category (other manufactures) are apparently associated with small workshops, we calibrate their weight E_i on the least-polluting category (wood processing).

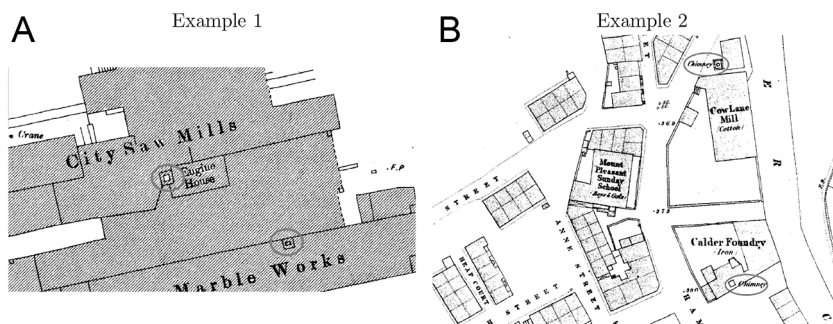A   Example 1   B   Example 2

FIG. 2.—Examples of markings used to identify chimneys in 1842–1952 Ordnance Survey maps (scaled at 25 inches to the mile). Four different symbols for chimneys are circled. The variation in symbols prevents us from directly using a recognition algorithm. Instead, we go through all maps and mark chimneys with the symbol X and a unique numeric identifier (see fig. 3).

2.   Pollution Dispersion

The previous exercise produces a map of industrial pollution sources. The ADMS 5 dispersion model then permits us to construct a citywide map of exposure to air pollutants (SO$_2$, measured in μg/m^3).[12] The ADMS 5 models atmospheric dispersion under a large spectrum of meteorological conditions, provides pollution estimates in coastal areas, incorporates the impact of temperature and humidity, and accounts for complex terrain and changes in surface roughness.

  The ADMS 5 model requires a number of inputs. First, it uses meteorological information. We use contemporary, 10-year statistical meteorological data as provided by the Met Office, thereby neglecting small changes in prevailing winds related to climate fluctuations between the nineteenth century and today. Figure 4 illustrates the wind provenance and intensity for two of the four regional models: Northern England and Southern England. Winds blow mostly from the west/southwest; it is, however, less predictable in Northern England, generating, on average, more dispersed air-pollution measures. Second, the model requires complex terrain data and convective meteorological conditions on land. We use the current terrain height and ruggedness, which affect wind speed and turbulence for cities with high gradients.[13] Finally, ADMS 5 requires information on the emission source. Atmospheric dispersion modeling is usually parameterized

---

[12] See http://www.cerc.co.uk/environmental-software/ADMS-model.html. Atmospheric dispersion models are additive, such that concentration of air pollutants is calculated as the sum of concentrations computed separately from each chimney.

[13] Since industrial chimneys during the Industrial Revolution were shorter than modern chimneys, pollution dispersion was heavily influenced by surrounding topography. In fig. A1, we show the differences in pollutant dispersion implied by topography in a city with high gradients (Oldham). Topography and land cover play little role in flat terrains.
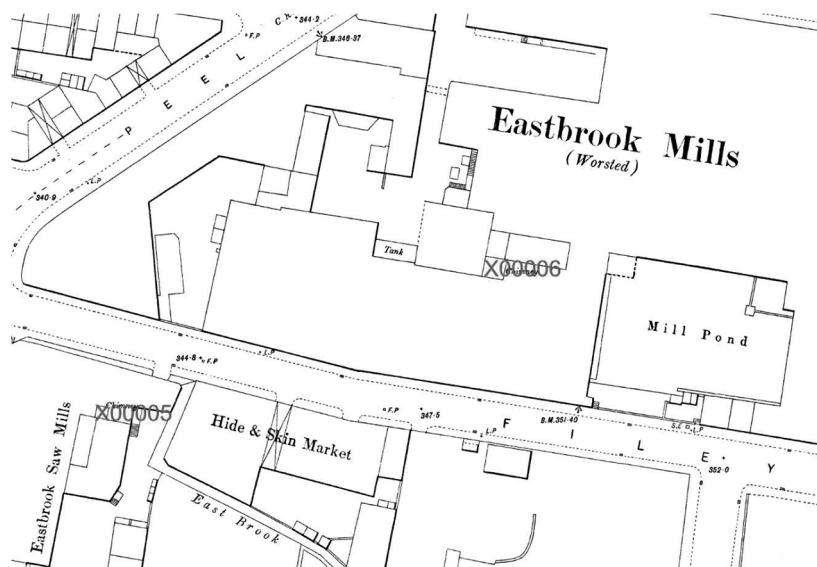
F𝚒ɢ. 3.—Marks X and identifiers (e.g., "00006") used by a recognition algorithm to locate chimneys and associated factories, together with the projection provided by the Ordnance Survey, allow us to geolocate each chimney. Factory-specific information can be retrieved after the recognition algorithm has (i) located a chimney and (ii) stored the associated identifier.

on current chimneys that are tall and wide and have high exit velocity. By contrast, chimneys in the Industrial Revolution were 10–50 m tall, with the majority being shorter than 25 m. Moreover, the exit velocity and temperature were also lower than today. To incorporate these characteristics, we set chimney height to 25 m in the baseline and assume an exit velocity of 4 m/s and an exit temperature of 120°C. To model pollution from residential sources, we assume domestic chimneys to be uniformly distributed within city borders at a very low altitude, and the ADMS 5 model is used under the same meteorological and topographic inputs.[14]

To validate our pollution measure, we use a sample of deposits collected in a few neighborhoods of Manchester as part of the First Annual Report of the Sanitary Committee on the Work of the Air Pollution Advisory Board, 1915 (Mosley 2013). We provide a comparison of our constructed measure with this external source in figure 5. We observe a large variation across neighborhoods for both measures, illustrating that distance to chimneys, topography, and wind directions generate significant

[14] In our sample of industrialized cities, the relative contribution of domestic emissions (vs. industrial emissions) in explaining the distribution of pollutant concentration within a city is low: while industrial coal consumption was very high but also very concentrated in few neighborhoods, residential coal consumption was equally spread across neighborhoods.
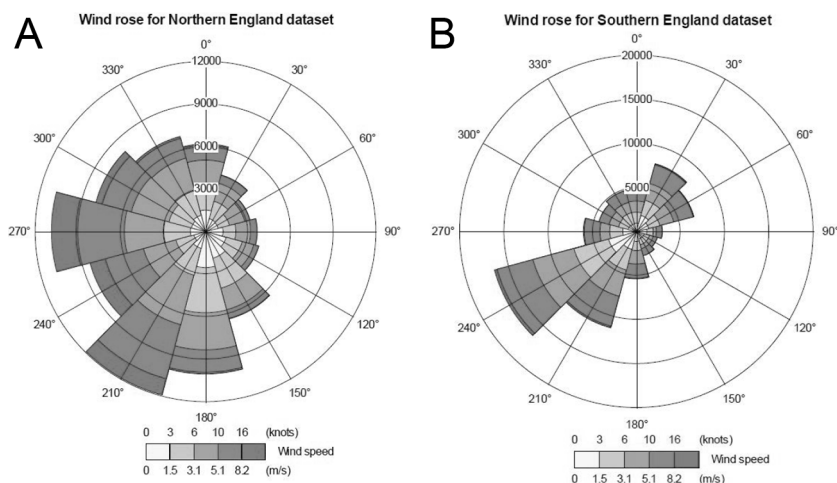
Fig. 4.—Wind-rose differences across two sets of meteorological conditions based on 10-year statistical meteorological data from the Met Office. *A*, North England. *B*, South England. We use four different sets of meteorological conditions across England and Wales: Southern England, Central England, Northern England, and East Anglia.

within-city dispersion in pollution, with some neighborhoods reaching alarming concentrations in air pollutants.[15] Reassuringly, the estimated pollution very strongly correlates with the deposit measure.

Another validation exercise brings us back to the example of the peppered moth and the appearance of the darker *carbonaria* form. We exploit a collection of surveys reporting the melanic forms of species of moths (Cook 2018). We restrict the sample to 54 surveys conducted before the swift decrease in industrial pollution (between 1965 and 1974). Figure A2 (figs. A1–A16 are available online) illustrates the relationship between our measure of historical pollution at each survey site and the share of darker moths. We find higher shares of the darker form (*carbonaria*) in highly polluted areas. There is a difference of about 60 percentage points between the least- and most-polluted survey sites. This exercise provides additional support for the validity of our pollution measure.

3. Measure of Neighborhood Composition

Measures of neighborhood sorting in the nineteenth century are extracted from individual records of the 1881 census. These records hold

[15] To better understand the extent to which cities were polluted at the end of the nineteenth century, we provide the cumulative distribution for our measure of pollution in our sample of neighborhoods (LSOA). Figure A3 shows that about 10% of LSOAs display air pollution above the two National Ambient Air Quality Standards (SO_2 concentration above 12 and 15 $\mu g/m^3$, respectively). About 2% of LSOAs—mostly in Manchester, Oldham, and Liverpool—have indexes of pollution above the peaks recorded in contemporary Beijing (40 $\mu g/m^3$).
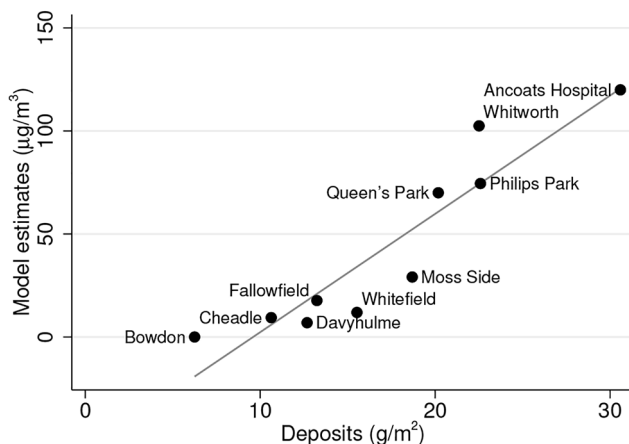
FIG. 5.—Air-pollution measures (external validity) across neighborhoods of Manchester, based on the First Annual Report of the Sanitary Committee on the Work of the Air Pollution Advisory Board, 1915. The graph charts the relationship between deposits, as collected by the Air Pollution Advisory Board (1915), and our measure of SO_2 concentration ($\mu g/m^3$). The correlation between the two measures is 0.92. Deposit measure (g/m^2) is available for the following neighborhoods: Ancoats Hospital (30.59), Philips Park (22.59), Whitworth Street (22.51), Queen's Park (20.18), Moss Side (18.69), Whitefield (15.53), Fallowfield (13.24), Davyhulme (12.68), Cheadle (10.63), and Bowdon (6.25).

information on the structure of households and each occupant's gender, age, occupation, and place of birth. There are two indicators of household location: a parish variable and an unreferenced address. While the parish variable is consistently referenced, the geographically more precise address information is inconsistently reported (surveyors use abbreviations, and misspelling is frequent) and poorly digitized (e.g., due to handwriting). To process this patchy information, we develop a method to allocate households interviewed in the 1881 census to small administrative units (2001 lower layer super output area [LSOA]), based on the organization of data collection.

Individual surveyors were given blocks to survey, and each filled in enumerator books while visiting their allocated neighborhoods: this induces a mechanical spatial clustering among adjacent individual records. The exact position of each entry in the 1881 census is thus an exceptional source of information that has, to the best of our knowledge, not previously been exploited.[16] If we locate a fraction of households, we can infer the location of unmatched entries given their position in the census books and the location of well-matched neighbors.

---

[16] Logan and Parman (2017) exploit the structure of the 1880 US census enumeration to create segregation measures based on the race of "census neighbors."

To implement our clustering analysis, we first geolocate a non-negligible fraction of households. First, we create a pool of geolocated addresses, heritage sites, and listed buildings. Second, we run a fuzzy matching procedure between census addresses and the pool of geolocated addresses within the same registration parish. A perfect match is found for 20% of all records, and a further 30% are matched with sufficient precision (where 90% of the original string is found in the matched address).[17] To locate the remaining addresses, we run an algorithm that uses the well-matched addresses together with the mechanical clustering induced by a surveyor's sequenced record taking to infer a location for the unmatched addresses. This procedure is described in detail in appendix D.

For the preperiod, we use "The Occupational Structure of England and Wales, c.1817–1881" (Shaw-Taylor and Wrigley 2014), which constructs a quasi-census of male occupations around 1817 using baptism records. These data are nested within the 834 parishes of the 1881 microcensus and cannot be allocated to smaller administrative units. Recent censuses (1971–2011) provide consistent measures of occupation, housing, education level, and country of origin, and we use area weights to map census enumeration districts into LSOAs.

One drawback is that we do not directly observe income. Instead, we observe three-digit occupational information in recent censuses and rely on a similar classification for 1817 and 1881 (the primary, secondary, tertiary [PST] system of classifying occupations; see Wrigley 2010). There are various ways to proxy for income based on occupational structure, for example by predicting income using average occupational wages. Such inference would require assumptions regarding the relative wage per occupation across cities. For the sake of transparency, we rely on a proxy based on the raw data, that is, the share of low-skilled workers among the working population. For 1817 and 1881, we collapse the 500 occupational subcategories into 10 categories. We restrict the sample to individuals with the lowest possible measurement error, that is, males between 25 and 55.[18] Unemployed, disabled, unskilled, and semiskilled workers are classified as low-skilled workers. Managers, gentlemen, rentiers, clerks, and manual skilled workers are classified as high-skilled workers. We assign farmers to a separate category and drop soldiers from our analysis. This breaks down to about 60% low-skilled workers, 30% high-skilled workers, and 10% farmers in 1881 (78%, 12%, and 10%, respectively, in 1817)

---

[17] There are three potential sources of noise when matching historical address with current addresses: (i) reporting error from past surveyors, (ii) digitizing errors, and (iii) changes in street names, e.g., red-light districts. The first two sources of error are the most common.

[18] Our results are robust to (i) adding female workers, as we will show later, and (ii) widening the age interval (e.g., 15–65).

in the 70 metropolitan areas. For 1971–2011, occupations are already classified into one-digit occupational categories: managers, professionals, associate professionals, administration, manual skilled, care, sales, processing, and elementary. We group the first three categories as high-skilled and the remaining six as low-skilled to harmonize shares of low-skilled workers between 1881 and 1971–2011. Clerks and manual skilled workers are thus classified as low-skilled, which breaks down to 62% low-skilled workers and 38% high-skilled workers in 2011.

C.   Descriptive Statistics

We start by providing evidence on the correlation between exposure to air pollutants and neighborhood composition and the underlying role of prevailing wind patterns. Figure 6 displays the spatial gradients in pollution and in the share of low-skilled workers at the end of the nineteenth century in the average city of our sample.[19] The pollution cloud leans toward the east, which could be due to prevailing winds or possibly the unequal distribution of pollution sources across space. The spatial gradient in the share of low-skilled workers also exhibits a similar asymmetric pattern toward eastern neighborhoods, albeit with slightly greater noise.

Figure 7 refines our analysis of the relationship between the share of low-skilled workers in 1881 and pollution sources. Units of observation are the 675,000 block × chimney pairs, where a block is a census cluster of households with the same geolocation in 1881 that is located within 2 km of the chimney (there are 100,000 such blocks). Figure 7A displays the average share of low-skilled workers in 1881 as a function of distance to the chimney. There is a sharp gradient, with a 10 percentage point difference between 100 and 1,500 m from a pollution source. This gradient likely captures high commuting costs. Strikingly, even conditional on distance to the pollution source and amenities in 1881, there remains large variation in the share of low-skilled workers at the block level. Part of this variation relates to the location of the block relative to the chimney. As apparent in figure 7B, there is a 1.5 percentage point excess share of low-skilled workers for blocks situated northeast of the chimney. This gradient in the direction of prevailing winds is what will be captured in the reduced-form analysis in the next section.

Table 1 provides summary statistics at the level of our baseline units of observation. Within a buffer of 20 km around the centroids of our 70

---

[19] Figure 6A is constructed as follows. First, we define a grid of equally spaced points—every 100 m—within 1.5 km of each city centroid, and we associate excess pollution at each point relative to the average city pollution. Second, we overlay the city grids and compute, for each point, the unweighted average of excess pollution across cities. Third, we interpolate across grid points using a Gaussian kernel interpolation method. The 20 level lines are quantiles of pollution. Figure 6B is constructed in the same fashion, with the share of low-skilled workers in 1881. We provide a similar illustration centered on town halls in fig. A4.
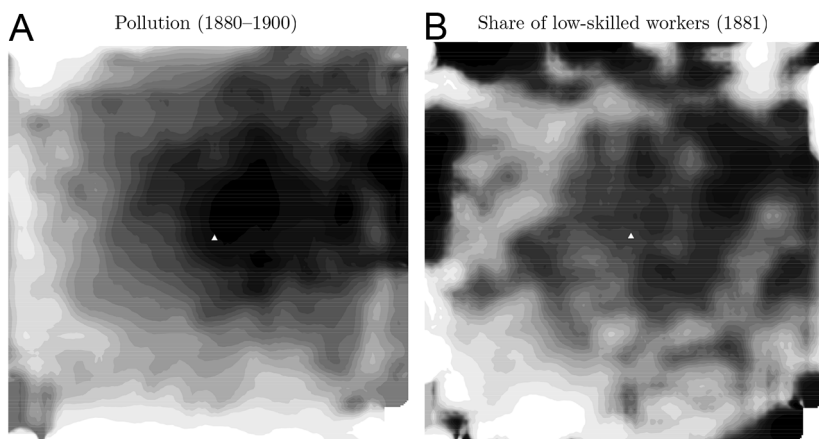
FIG. 6.—Pollution and share of low-skilled workers in the average city. *A*, Average gradient of pollution in 1880–1900 across cities. *B*, Average gradient of neighborhood composition, as captured by the share of low-skilled workers in 1881, across cities. To construct this figure, we define a grid of equally spaced points within 1.5 km of each city centroid (indicated with a triangle); we overlay the grids across cities and consider the unweighted average. In order to create a continuous measure, we interpolate across the grid points using a Gaussian kernel. The 20 shadings/level lines are quantiles of pollution (*A*) and share of low-skilled workers (*B*); darker shadings indicate the highest quantiles.
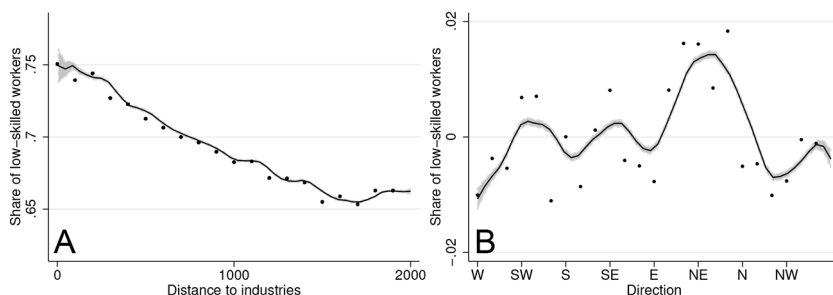
FIG. 7.—Relationship between the share of low-skilled workers in 1881 and their position relative to a pollution source. In this exercise, the unit of observation is a neighborhood × chimney pair, where a neighborhood is a census cluster of households with the same geolocation in 1881 (about 100 households). *A*, Average share of low-skilled workers in 1881 across observed units (weighted such that all households are given the same weight). *B*, Residual of the share of low-skilled workers in 1881 cleaned for distance to the pollution source and distance to amenities (canals, town hall, theaters, hospitals, parks, churches, schools, universities, guild hall, mills, and elevation), as a function of the direction with respect to the pollution source. The lines are locally weighted regressions on all observations with respective bandwidths of 30 m (*A*) and 20° (*B*). "NE" indicates that the household is located toward the northeast direction, from the standpoint of the pollution source.

TABLE 1
Descriptive Statistics and Variance Decomposition

	Mean (1)	Pollution High (2)	Low (3)	Standard Deviation Total (4)	Between (5)	Within (6)
			A. Air Pollution			
Normalized pollution	−.034	.234	−.298	.928	.596	.560
			B. Population Measures			
1817:[a]						
Low-skilled workers	.784	.787	.780	.111	.090	.073
High-skilled workers	.093	.085	.102	.087	.085	.068
Farmers	.123	.128	.117	.092	.066	.047
Property tax (log)	9.94	10.05	9.82	1.25	1.05	.077
1881:						
Low-skilled workers	.601	.627	.573	.256	.149	.235
High-skilled workers	.278	.285	.271	.241	.125	.217
Farmers	.121	.088	.156	.199	.176	.180
2011:						
Low-skilled workers	.587	.605	.568	.173	.118	.123
High-skilled workers	.413	.395	.432	.173	.118	.123
			C. Topography Controls			
Maximum elevation (m)	72.9	66.9	79.0	66.0	63.3	31.8
Minimum elevation (m)	52.6	50.2	54.9	48.9	44.1	19.6
Mean elevation (m)	62.3	58.3	66.4	55.8	51.6	23.4
Distance canals (km)	6.12	5.57	6.69	14.9	19.1	1.30
			D. Amenities Controls			
Distance town hall (km)	4.64	3.05	5.24	5.35	4.72	1.27
Distance parks (km)	9.57	9.38	9.77	23.9	28.9	1.16
Share LSOA within city	.296	.398	.192	.418	.245	.296
Area (km²)	.940	.608	1.28	3.99	8.00	3.50
Distance heavy (km)	2.49	1.98	3.01	6.36	8.82	.973
Distance light (km)	5.32	4.80	5.84	13.5	17.6	1.16

Note.—All statistics are computed using the baseline sample of 5,538 lower layer super output areas (LSOAs). Standard deviations are decomposed into between- and within-city standard deviations. The samples of high and low within-city pollution are defined with respect to the median city pollution.

[a] Computed at the parish level, which explains the lower variance.

metropolitan areas, the clustering process associates about 5 million active male workers in 1881 to 5,538 LSOAs. As these LSOAs are the 2001 census units, we can associate contemporary measures to all of these 5,538 geographic units that will constitute our baseline sample. The sample covers 70 metropolitan areas, 142 "cities" (i.e., local administrations), and 542 parishes.[20] We provide summary statistics for the full sample and for LSOAs with above- and below-median pollution at the city level. We report statistics for

[20] We show the geographic distribution of these 5,538 LSOAs in fig. A10.

the main outcome variables and baseline controls, accounting for topography, amenities, and direction (latitude and longitude). Some of these characteristics capture important differences between more- and less-polluted LSOAs within cities. Less-polluted neighborhoods have higher elevation, are more rural, and are more distant from waterways and pollution sources. In columns 4–6 of table 1, we provide a decomposition of the variance within and between cities. A very large share of the variance in pollution is within cities. Our empirical strategy, described in the following section, hinges on such within-city variation and is mostly orthogonal to variation across cities.

III.   Reduced-Form Evidence

This section presents reduced-form evidence on (i) historical pollution and neighborhood sorting and (ii) the subsequent persistence of neighborhood segregation.

A.   Empirical Strategy

To estimate the impact of pollution on neighborhood sorting within cities, we run a baseline difference specification at the LSOA level and an IV specification where we employ exogenous factors that influence the location of pollution sources.

1.   Baseline Specification

Letting i denote an LSOA, p a parish, c a city, and t a particular census wave, we estimate the following equation:

$$Y_{it} = \alpha + \beta P_i + \gamma \mathbf{X}_i + \mathbf{v}\mathbf{Y}_p + \delta_c + \varepsilon_{ict}, \qquad (S1)$$

where Y_{it} is a measure of occupational structure. The measure of historical pollution, P_i, results from a combination of the location of pollution sources and a dispersion process. Physical features such as hills or rivers that enter the simulated pollution measure may also be local (dis)amenities that affect individual neighborhood choices. To eliminate this potential source of bias, we include separate topography indicators (e.g., maximum, minimum, and average elevation), along with a rich set of geographic controls (e.g., area, share of LSOA within the city borders, latitude, and longitude) and controls for (dis)amenities (distance to waterways, heavy-industry chimneys, light-industry chimneys, the town hall, and parks) in the set of controls $\mathbf{X}_i$. We use $\mathbf{Y}_p$ to indicate a set of measures of occupational structure in 1817 at the parish level (shares of low-skilled workers, high-skilled workers, and farmers) and the logarithm of the

property tax in 1815 at the parish level to capture possible fixed neighborhood amenities. City fixed effects are indicated by δ_c, and standard errors are clustered at the parish level.[21]

We further exploit the interaction between the distribution of pollution sources and air-pollution dispersion by considering counterfactual diffusion processes, such as those generated by the same pollution sources but with artificially rotated wind patterns. One can think of this procedure as a decomposition of the interaction between location and diffusion; it isolates variation induced by the asymmetry between neighborhoods at the same distance from factories, some of them being located downwind and others upwind (as in fig. 7).

A concern with specification (S1) is that the treatment may not be exogenous because fixed unobserved amenities explain both the upwind presence of industries and the local occupational structure. In robustness checks, we show a balance test before the rise of industrial coal pollution and provide identification at a more granular level by including fixed effects at the level of parishes or electoral ward.[22]

Finally, there is a remaining threat to identification from reverse causality or time-varying omitted variation. For instance, factories may be strategically placed upwind of poor neighborhoods to minimize political or economic costs associated with environmental disamenities in richer neighborhoods. We address this concern with an instrumental variable.

2. IV Specification

To account for the bias arising from potentially nonrandom industry location, we exploit exogenous variation in location factors, which translates into exogenous variation in pollution imprints. Specifically, we exploit the fact that large boilers required a constant stream of water for cooling. As a result, the natural geographic placement of all mills was along rivers or canals (Maw, Wyke, and Kidd 2012). We locate hypothetical chimneys in intervals of 150 m along waterways in 1827, before the rise of coal as the

---

[21] Our main independent variable, i.e., historical pollution, is a preestimated regressor, which has possibly a sampling variance of its own, due to, e.g., variation in the 10-year average weather conditions. Standard errors need to be adjusted in such cases (Murphy and Topel 2002). One standard method is to bootstrap the procedure defined by (i) the estimation of historical pollution and (ii) our reduced-form regression. However, the sampling variance generated by the first step occurs within the atmospheric dispersion model that we outsource to ADMS 5. We thus cannot correct our standard errors for this issue, and they may be slightly underestimated. On a separate note, we allow for spatial correlation along distance in robustness checks (following Conley 1999), instead of clustering standard errors at a certain administrative level.

[22] One issue with our reduced-form approach is that the potential outcome for one neighborhood is affected by the treatment intensity in other neighborhoods. Finer fixed effects may aggravate this issue if individuals choose residences within a small radius around their working place. One way to deal with this issue is to develop a proper model of neighborhood choice that accounts for equilibrium adjustments at the city level (see sec. IV).

main energy source. To derive our instrument, we assume uniform air-pollutant emissions from these exogenous pollution sources, combined with the actual atmospheric dispersion due to wind flows and topography.[23] This natural geographic placement of chimneys is not susceptible to being selectively placed upwind of poor neighborhoods. However, the variation correlates with proximity to waterways that may itself affect the attractiveness of a neighborhood. We thus control separately for distance to waterways and exclude neighborhoods bordering these waterways.

We then use the following first-stage specification to instrument the historical pollution, P_i, in specification (S1):

$$P_i = b_0 + b_1 PP_i + \mathbf{c}\mathbf{X}_i + d_c + \mathbf{f}\mathbf{Y}_p + e_{ict}, \qquad (S2)$$

where PP_i is the simulated pollution using hypothetical chimneys. As described above, $\mathbf{X}_i$ includes a comprehensive set of controls for physical attributes, $\mathbf{Y}_p$ is the occupational structure in 1817 at the parish level, and d_c are city fixed effects.

B. Historical Pollution and Neighborhood Sorting

In this section, we document a positive contemporaneous correlation between air pollution and the share of high-skilled workers in 1881. In table 2, we report the estimates for our baseline specification (S1). As can be seen in column 1, air pollution and the share of low-skilled workers in 1881 are positively correlated. It does not affect the estimates when we control for a large set of covariates. In column 2, we add city fixed effects to control for variation in atmospheric pollution and neighborhood composition between cities (Hanlon 2019). In column 3, we add (log) property tax in 1815 and the parish-level shares of low-skilled workers, high-skilled workers, and farmers in 1817 to clean for potentially unobserved fixed characteristics. For columns 4–6, we add separate elements entering in the pollution-dispersion process. In column 4, we condition on topography (elevation and distance to waterways in 1827). In column 5, we control for distance to pollution sources (heavy and light industry), distance to city hall, distance to parks, area, and the share of the LSOA within the 1880 city borders.[24] In column 6, we add eastings

---

[23] Figure A5 depicts our approach. In panel *a*, we see the cities of Manchester and Oldham with the associated 1827 natural waterways. Panel *b* displays the natural geographic placement of chimneys along canals, and panel *c* shows the resulting spatial distribution of air pollutants. Finally, panel *d* shows the distribution of air pollutants using actual pollution sources.

[24] As stated in sec. II, the metropolitan areas (and thus the sample of LSOAs) are defined by a buffer of 20 km around the centroids of our cities. In robustness checks, we verify that the results are left unchanged if we limit the sample to urban LSOAs intersecting with the 1880 city borders (which may be endogenous and affected by pollution through agricultural yields).

TABLE 2
POLLUTION AND SHARES OF LOW-SKILLED WORKERS IN 1881

	(1)	(2)	(3)	(4)	(5)	(6)
Pollution	.0440	.0428	.0409	.0378	.0354	.0338
	(.0071)	(.0073)	(.0068)	(.0066)	(.0067)	(.0069)
	[.1738]	[.1690]	[.1614]	[.1491]	[.1398]	[.1332]
Observations	5,538	5,538	5,538	5,538	5,538	5,538
Fixed effects (city)	No	Yes	Yes	Yes	Yes	Yes
Controls:						
Population	No	No	Yes	Yes	Yes	Yes
Topography	No	No	No	Yes	Yes	Yes
Amenities	No	No	No	No	Yes	Yes
Latitude/longitude	No	No	No	No	No	Yes

NOTE.—Standard errors are reported in parentheses and clustered at the parish level. Standardized effects are reported in brackets. Each cell is the result of a separate regression. The unit of observation is a lower layer super output area (LSOA). The set of population controls include the parish-level shares of farmers, managers, and blue-collar workers in 1817; the logarithm of the average property tax at the parish level in 1815; and total population in 1881. The set of topography controls include the average, maximum, and minimum elevations for the LSOA and the (inverse) distance to waterways as of 1827. The set of amenities controls include the (inverse) distance to the city hall, (inverse) distance to parks, share of LSOA within the city borders in 1880, LSOA area, (inverse) distance to the closest heavy industry, and (inverse) distance to the closest light industry. Latitude/longitude are for the LSOA centroid.

and northings of the LSOA centroids to control for wind patterns and potential western or southern preferences in locations. As apparent from table 2, our estimates slightly decrease but remain large and precisely estimated.[25]

The correlation between air pollution and the occupational structure is both statistically and economically significant. In the baseline specification (col. 6), the coefficient is 0.034 and the 95% confidence interval is [0.020, 0.047]. One additional standard deviation in air pollution increases the prevalence of low-skilled workers by 3.4 percentage points, which is about 13.3% of a standard deviation in their prevalence across LSOAs. A differential in pollution equivalent to the one between the first and last deciles in Manchester would be associated with a differential of 16 percentage points in the share of low-skilled workers.[26] Residential sorting may be tempered by the necessity for residents to live close to their working place, as induced by relatively costly modes of transportation.

Figure 8 illustrates the estimated relationship between the share of low-skilled workers before and after the rise in coal use (in 1817 and 1881, respectively) and the atmospheric pollution during the Industrial Revolution. On the *y*-axis, we plot the residuals from a regression of the standardized shares of low-skilled workers on a similar set of controls, as in column 6 of table 2. On the *x*-axis, we plot the regression-adjusted residual of standardized air pollution. The relationship between the share of low-skilled workers and standardized air pollution is strongly positive but flattens at both extremes, that is, for very high and very low within-city pollution levels. By contrast, there is no correlation between the share of low-skilled workers in 1817 and the measure of atmospheric pollution.

The most convincing evidence in support of a causal relationship between pollution and neighborhood sorting comes from hypothetical pollution imprints. In figure 9, we disentangle the role of prevailing winds from the correlation induced by proximity to pollution sources.[27] We first construct a measure capturing average proximity to pollution sources at the LSOA level. The measure of symmetric pollution is generated by dispersing pollutants from existing chimneys but under a wind profile that is

---

[25] Column 1 of table A2 reports the coefficients on all covariates.

[26] We consider other outcomes in table A3, with the share of all low-skilled workers including females and the share of migrants distinguishing between migrants from England and Wales vs. the Commonwealth. We find that the standardized effects of pollution on the share of all low-skilled workers and migrants are comparable to the baseline findings. Interestingly, the higher prevalence of migrants in polluted neighborhoods is essentially due to migrants from England and Wales (and thus unrelated to the Irish Potato Famine).

[27] The exercise underlying fig. 9 is similar in nature to the exercise underlying fig. 7. In both cases, the objective is to exploit the direction of prevailing winds and the location of neighborhoods relative to chimneys, with controls for their proximity to such chimneys. Relative to fig. 7, fig. 9 isolates the role of prevailing winds (i) by aggregating pollution exposure across chimneys and census blocks and collapsing the data at the LSOA level and (ii) by conditioning the analysis on an extended set of controls (as in col. 6 of table 2).
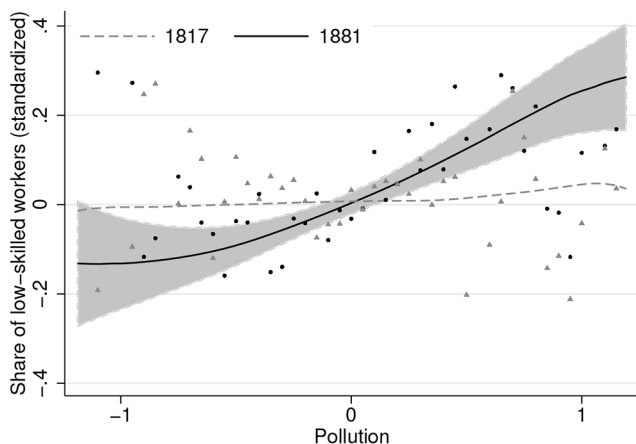
Fig. 8.—Pollution (*x*-axis) across neighborhoods and shares of low-skilled workers (*y*-axis) in 1817 and 1881. The graph charts the relationship between the (standardized) shares of low-skilled workers in 1817 (gray triangles) and 1881 (black diamonds) and our (standardized) measure of past pollution. We consider the residuals of all measures once cleaned by the topography controls, the amenities controls, and the latitude/longitude controls (see table 1). We create 40 bins of neighborhoods along past pollution, and the circles represent the average shares of low-skilled workers within each bin. The lines are locally weighted regressions on all observations. We restrict the sample to observations with residual pollution between −1 and 1 standard deviation.

symmetric in all directions. We then generate a set of counterfactual pollution exposures using wind profiles rotated in steps of 30° relative to the actual prevailing winds. Figure 9 displays the correlations between these rotated measures and the share of low-skilled workers for the years 1817 and 1881, conditioning on an extended set of controls and the symmetric pollution measure. In 1817 (fig. 9*A*), before the rise of coal pollution, we observe virtually no correlation between rotated pollution measures and the share of low-skilled workers. In 1881 (fig. 9*B*), after pollution became a meaningful disamenity, we see a pronounced bell-shaped pattern, with a peak in correlation observed around wind profiles rotated by 0° and 30°.[28] As we rotate wind profiles away from prevailing winds, the estimated relationship loses significance and turns negative. To reduce measurement error, we clean our estimates for parish fixed effects in figure 9*C* and for residential pollution in figure 9*D*. The estimates remain large within a narrow corridor along prevailing winds, but they now decrease sharply, becoming negative for rotations of more than 90°.

[28] The fact that the peak in correlation is 0°–30° may be due to measurement error. First, wind patterns may have changed in one century, in particular the frequency of cyclonic or anticyclonic conditions (Lamb 1972), each associated with different wind-direction profiles. Second, we consider yearly averages for our meteorological conditions, possibly ignoring differential pollution exposure and wind patterns across seasons or hours of a day.
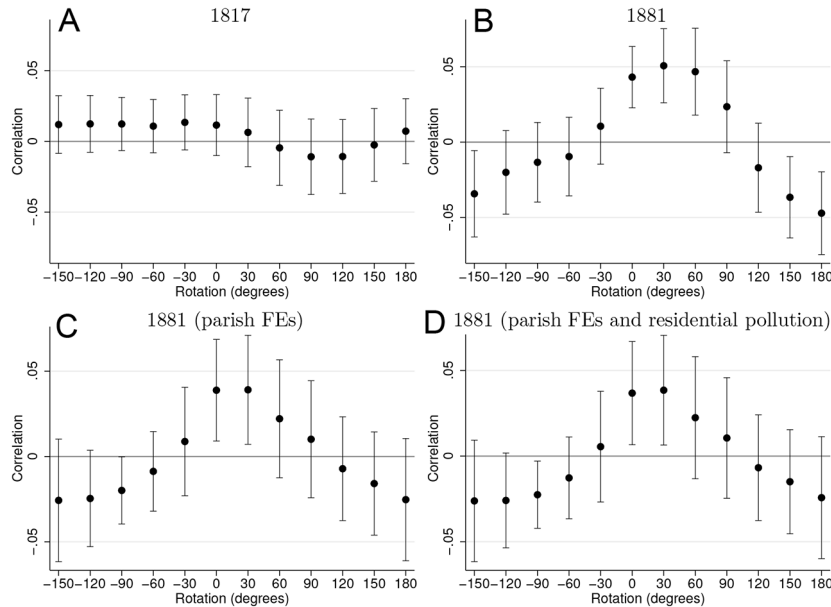
FIG. 9.—Rotating wind patterns representing conditional correlations between shares of low-skilled workers and counterfactual measures of past pollution rotated in steps of 30°. *A*, 1817; *B*, 1881; *C*, 1881, with controls for parish fixed effects; *D*, 1881, with controls for parish fixed effects (FEs) and residential pollution. Each circle represents the estimate in a specification including the controls reported in table 1, column 6, and the measure of symmetric pollution capturing proximity to the pollution source. Standard errors are clustered at the parish level, and the lines represent 5% confidence intervals.

Finally, we present in table 3 the results of the IV strategy (spec. [S2]), that uses 1827 waterways as a source of exogenous variation for chimney locations. We report four sets of estimates, excluding neighborhoods within 250 or 500 m of waterways, both with and without extended controls in addition to city fixed effects. The first stage is strong in all cases, and the (stable) two-stage least squares estimates tend to be larger than the OLS estimates. One additional standard deviation in air pollution increases the prevalence of low-skilled workers by about 9 percentage points. Explanations for the downward bias in the OLS specification could be measurement error or possibly the difference between the local average treatment and the average treatment effects. The IV strategy mostly relies on variation induced by pollution sources located around city centers, and the treatment effect appears larger for these central neighborhoods. Indeed, the local slope between pollution and the share of low-skilled workers is highest for pollution levels between 0 and 0.50 SD above the mean (see fig. 8); the average pollution for central neighborhoods or for neighborhoods in proximity to waterways is around

TABLE 3
POLLUTION AND SHARES OF LOW-SKILLED WORKERS IN 1881—INSTRUMENTAL
VARIABLES SPECIFICATION

	(1)	(2)	(3)	(4)
	A. First Stage: Pollution			
Pollution (waterways)	.2904	.2020	.3100	.1834
	(.0331)	(.0347)	(.0403)	(.0384)
	B. Second Stage: Share of Low-Skilled Workers			
Pollution	.1286	.0937	.1143	.0695
	(.0199)	(.0296)	(.0201)	(.0359)
	[.5076]	[.3695]	[.4511]	[.2743]
Observations	4,830	4,830	4,557	4,557
F-statistic	77.16	33.91	59.22	22.81
OLS coefficient	.0408	.0216	.0392	.0209
Sample	Canal > 250 m	Canal > 250 m	Canal > 500 m	Canal > 500 m
Fixed effects (city)	Yes	Yes	Yes	Yes
Extended controls	No	Yes	No	Yes

NOTE.—Standard errors are reported in parentheses and clustered at the parish level. Standardized effects are reported in brackets. Panel A reports the first stage, and Kleibergen-Paap F-statistics are reported in panel B. The unit of observation is a lower layer super output area (LSOA). The set of extended controls include all controls in col. 6 of table 2. The variable pollution (waterways) is the first predicted pollution instrument from a uniform allocation of pollution sources along waterways (as of 1827). Samples exclude LSOAs within 250 m (cols. 1, 2) and 500 m (cols. 3, 4) of a waterway.

0.10–0.30. Moreover, we find larger average treatment effects when restricting the sample to central LSOAs (see app. E).

C. Historical Pollution and Contemporary Neighborhood Segregation

This section extends our analysis of neighborhood sorting to recent census waves (1971–2011) to assess potential reversion to the mean after the 1968 Clean Air Act stopped coal use within cities. Table 4 reports the slopes between the shares of low-skilled workers and historical pollution,

TABLE 4
POLLUTION AND SHARES OF LOW-SKILLED WORKERS IN 1971–2011

	1971	1981	1991	2001	2011
Pollution	.0244	.0309	.0388	.0374	.0355
	(.0046)	(.0050)	(.0063)	(.0063)	(.0057)
	[.1914]	[.2204]	[.2072]	[.2299]	[.2029]
Observations	5,535	5,538	5,538	5,538	5,538
Fixed effects (city)	Yes	Yes	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes	Yes	Yes

NOTE.—Standard errors are reported in parentheses and clustered at the parish level (as defined in 1881). Standardized effects are in brackets. Each cell is the result of a separate regression. The unit of observation is a lower layer super output area. The set of extended controls include all controls in col. 6 of table 2.

TABLE 5
POLLUTION AND SHARES OF LOW-SKILLED WORKERS IN 1971–2011—INSTRUMENTAL
VARIABLES SPECIFICATION

	1971	1981	1991	2001	2011
Pollution	.0300	.0406	.0342	.0462	.0495
	(.0164)	(.0198)	(.0257)	(.0193)	(.0214)
	[.2358]	[.2895]	[.1823]	[.2839]	[.2830]
Observations	4,829	4,830	4,830	4,830	4,830
F-statistic (first stage)	33.91	33.91	33.91	33.91	33.91
Fixed effects (city)	Yes	Yes	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes	Yes	Yes

NOTE.—Standard errors are reported in parentheses and clustered at the parish level (as defined in 1881). Standardized effects are in brackets. Each cell is the result of a separate regression. The unit of observation is a lower layer super output area (LSOA). The set of extended controls include all controls in col. 6 of table 2. As in cols. 1 and 2 of table 3, the instrument is the predicted pollution generated by a uniform allocation of pollution sources along waterways (as of 1827), and we exclude LSOA within 250 m of a waterway.

as estimated by specification (S1). One standard deviation in historical air pollution increases the prevalence of low-skilled workers by 2.5–4 percentage points without a clear pattern between 1971 and 2011, and the standardized effects range between 0.19 and 0.23. Table 5 displays the IV estimates for the occupational structure in recent years and shows, as in table 4, that there are no signs of (overall) reversion to the mean.

Our analysis of spatial inequalities in cities of the nineteenth century was not informed by house or land prices, due to data scarcity. We do have such data for more recent years, however. In table 6, we use transactions

TABLE 6
POLLUTION, HOUSE PRICES, AND TRANSACTIONS (Nationwide, 2009–13,
and Land Registry, 2000–2011)

	NATIONWIDE		LAND REGISTRY	
	(1)	(2)	(3)	(4)
Pollution	−.1035	−.0852	−.1116	−.0642
	(.0168)	(.0121)	(.0161)	(.0121)
	[−.1685]	[−.1386]	[−.2030]	[−.1168]
Observations	5,226	5,226	5,538	5,538
Fixed effects (city)	Yes	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes	Yes
Controls (house characteristics)	No	Yes	No	Yes

NOTE.—Standard errors are reported in parentheses and clustered at the parish level. Standardized effects are in brackets. Each column is the result of a separate regression. The unit of observation is a lower layer super output area. The set of extended controls include all controls in col. 6 of table 2. The dependent variables are the (log) average house prices. The unreported effect of pollution on the number of transactions is between −0.0473 (specification similar to col. 3) and −0.0878 (specification similar to col. 4). In col. 2, controls for house characteristics include the average shares of new houses, the average square meters, number of bedrooms, and the year of construction for nationwide transactions. In col. 4, controls for house characteristics include the average shares of detached, semidetached, terraced, and new houses for all transactions.

in England and Wales as recorded by (i) the Land Registry between 2000 and 2011 (cols. 1, 2) and (ii) the Nationwide Building Society between 2009 and 2013 (cols. 3, 4). Hedonic regressions with and without controls for average house characteristics find that one additional standard deviation in past pollution is associated with a price drop of about 10%–11%. Controls for property characteristics reduce these estimates to 6%–8%, showing that properties in formerly polluted neighborhoods are smaller and more likely to be nondetached.[29]

Past environmental disamenities appear to have a marked effect on spatial inequalities today. A differential in pollution equivalent to the one between the first and last deciles in Manchester is associated with a differential of 16 percentage points in the share of low-skilled workers or with differences in property prices of about 40%.

To visualize possible nonlinearities in the persistence of neighborhood sorting, figure 10 displays the relationship between shares of low-skilled workers in 1817 (long dash), 1881 (short dash), 1971, 1991, and 2011 (plain lines) and the historical pollution disamenity that stopped after the 1968 Clean Air Act. As apparent, we observe some reversion to the mean for low and intermediate values of within-city pollution. By contrast, segregation patterns appear to persist at around 1 SD above average within-city pollution.

To take a closer look at the underlying dynamics, we refine the analysis between 1971 and 2011 and organize the data in a panel structure with decadal observations for each LSOA. To shed light on nonlinearities, we define 10 pollution categories corresponding to the 10 deciles in intracity pollution, that is, neighborhood pollution adjusted by the average city pollution. We then run a panel regression with the share of low-skilled workers as the dependent variable, LSOA fixed effects, city $\times$ year fixed effects, and trends for each pollution decile. The initial relationship between the share of low-skilled workers in 1971 and pollution deciles is displayed in figure 11A, while the estimates for pollution-decile trends are shown in figure 11B. A process of mean reversion would be captured in the form of a decreasing pattern in trends for each pollution decile.[30] Our findings are not consistent with a uniform reversion to the mean. Figure 11A shows evidence of mean reversion over the period 1971–2011 but only for areas with below-median levels of past pollution exposure (i.e., pollution categories 1–5). We see a slight increase in the share of low-income workers in the least-polluted neighborhoods and a corresponding decrease in moderately polluted neighborhoods. For

---

[29] Figure A6 illustrates these very large effects.

[30] Letting $x_{i,t}$ denote the excess share of low-skilled workers in neighborhood i and period t and $\theta < 1$ the AR(1) parameter, we should observe that

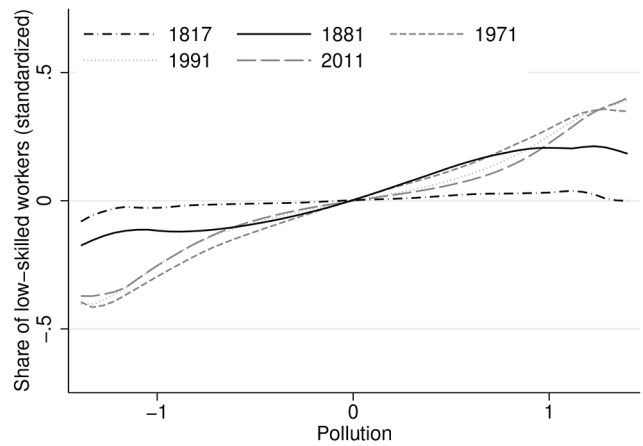$$E[x_{i,t+1} - x_{i,t}] = (\theta - 1)E[x_{i,t}].$$

FIG. 10.—Pollution (*x*-axis) across neighborhoods and shares of low-skilled workers (*y*-axis) in 1817, 1881, 1971, 1991, and 2011. This figure represents the locally weighted regressions on all observations between the (standardized) shares of low-skilled workers and our (standardized) measure of past pollution. We consider the residuals of all measures once cleaned by city fixed effects as well as topography and population controls.

neighborhoods with above-median levels of past pollution, however, we see the opposite. These neighborhoods become even more deprived relative to the median polluted neighborhoods. This pattern would be consistent with tipping dynamics leading to a high persistence of deprivation in neighborhoods with extreme pollution exposure in the past.
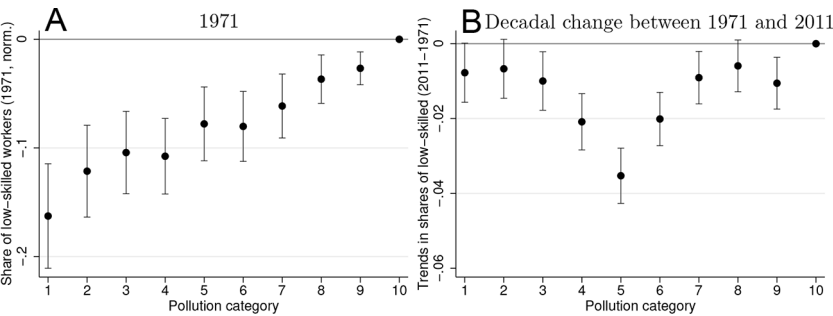
FIG. 11.—Mean reversion between 1971 and 2011—share of low-skilled workers by pollution decile. *A*, The effect of pollution on the share of low-skilled workers in 1971 in a specification with pollution category dummies defined by pollution decile (1 = lowest, 10 = highest pollution exposure). The effect of the last decile is normalized to 0. *B*, The effect of pollution on the annualized trends in low-skilled workers between 1971 and 2011. The specification is a panel regression using data in 1971, 1981, 1991, 2001, and 2011 with lower layer super output area fixed effects. The reported coefficients are extracted from the interaction of pollution category dummies and the year and are adjusted to represent decadal changes.

D.   Sensitivity and Robustness Checks

We conduct a large number of robustness checks around the baseline specification(s). We summarize the findings in this subsection and leave a detailed discussion of the results along with additional figures and tables to appendix E.

First, we conduct balance tests in the period before coal pollution to further reduce concerns about biasing effects from unobserved preexisting neighborhood characteristics. There is no correlation at the parish level between either the 1817 share of low-skilled workers or the 1815 property tax returns (as a proxy for wealth) and the later atmospheric pollution.

Second, we consider variations in our pollution modeling. We vary only the chimney height, as the exit velocity and temperature would affect the same crucial model input, that is, the height of the smoke column in the atmosphere. The atmospheric pollution based on chimneys assumed to be shorter (15 m) or taller (40 m) than our baseline (25 m) generates similar estimates, as in tables 2 and 4.

We also try a simple, albeit less informative, measure of pollution exposure. We overlay each city with a grid of equally spaced points and count the number of chimneys that are located in the (i) northeast, (ii) northwest, (iii) southeast, or (iv) southwest quadrants—within a given distance from each grid point. Last, we collapse the measure for all four quadrants into 2001 LSOAs. Only chimneys located in the northwest and southwest are correlated with the share of low-skilled workers, which coincides with prevailing downwind directions. Once we split the sample into northern and southern cities, we find that chimneys located in southwest quadrants are relatively more predictive of deprivation in northern cities, reflecting the more southerly wind direction in North England (see fig. 4).

Third, we test the sensitivity of our findings to the addition of the following pollution imprints as controls: a symmetric placebo pollution measure that assumes that pollution spreads evenly in all directions and captures the proximity to industrial centers; a placebo pollution pattern that varies the emission intensity by coding the chimneys of high (low) polluting industries as low (high) polluting; residential pollution; and a contemporary measure of atmospheric pollution. We find that the two placebo patterns do not affect the significance of the main pollution variable and have no predicting power, suggesting that there is no additional information in distance or emission intensity that might affect our estimation. Residential pollution does predict deprivation, but its standardized effect is one-third of the standardized effect of industrial pollution. Contemporary pollution has a small impact on neighborhood composition in 2011. Historical pollution is far more predictive of current spatial inequalities than current exposure to air pollutants.

Fourth, we explore sensitivity to controls, fixed effects, clustering, and sample selection. We control for a number of within-city geography variables interacted with city fixed effects to capture city-specific geographic patterns and commuting infrastructure: latitude and longitude, distance to the town hall, distance to heavy industries, and distance to light industries. The effect of a standard deviation in pollution remains stable at 3.6–4.5 percentage points in the share of low-skilled workers. We also report the results of our baseline specification with 540 parish fixed effects, with fixed effects expanded to 1,440 electoral wards and to the 1,850 middle layer super output areas (MSOA). The estimates remain unchanged, even when identification comes from a within-MSOA comparison. We report standard errors clustered at three different levels: electoral ward, MSOA, and city. Standard errors increase by about 40% between the least and most conservative choice, and our baseline analysis clustered at the parish level is at the center of this interval. Finally, we estimate the baseline specification on alternative samples where we exclude Greater London, the northwest, and the northeast, respectively. The estimates fluctuate around the baseline, but they remain large in all cases.[31] We also analyze the sensitivity of our results to the exclusion of suburbs and rural LSOAs. The estimates remain precisely estimated and larger than in the baseline when restricting the analysis to central neighborhoods.

Fifth, we consider an alternative instrument, based on the historical settlements of industries. We isolate variation induced by the location of industrial districts before coal became a major energy source that would affect disproportionately downwind neighborhoods. To predict early industrial districts, we locate 543 early steam engines installed between 1700 and 1800, using data from Kanefsky and Robey (1980) and Nuvolari, Verspagen, and von Tunzelmann (2011). We model uniform air-pollutant emissions from early steam engine locations and use the resulting atmospheric dispersion as an instrument for actual pollution, conditioning for distance to the nearest industrial chimneys. The estimates from using this instrument are remarkably similar to the ones using waterways as an exogenous factor for industry location.

Finally, we provide some insight on the recent dynamics of neighborhood composition (1971–2011) in appendix F, where we document differences across neighborhoods in school supply, crime, housing quality,

---

[31] We find, however, that the treatment effect is about three times smaller in the northwest region than in the rest of the country. The evidence suggests that it is due to nonlinearities in the treatment effect: pollution is about 0.8 SD above the country average in the northwest, and the relationship between pollution and deprivation flattens for such pollution levels in 1881 (see fig. 10). From 1971 to 2011, the relationship is more linear, and the northwest treatment effect is then much closer to the average treatment effect.

and public amenities. We also show how the liberalization of social housing and immigration inflows may have contributed to residential segregation between 1971 and 2011 in appendix G.

IV.   A Dynamic Model of Residential Sorting

In order to quantify the nonlinear dynamics in the persistence of neighborhood sorting between 1971 and 2011, we develop a dynamic model of residential sorting within a city, in which infinitely lived households face relocation frictions. Those relocation frictions mean that optimal residential choice depends on past and future neighborhood amenities. The purpose of the model is to derive a dynamic demand equation for neighborhoods, which we can then proceed to estimate in section V.

A.   Environment

Consider a unit mass of infinitely lived households, each of measure zero. Time is discrete, and each household receives utility every period from the consumption of the numeraire and a neighborhood amenity. Households are hand-to-mouth consumers, and they rent one unit of land from absentee landlords in a closed city. Land markets are competitive, and land supply is constant over time.

We assume that there is a discrete number of neighborhoods J. Let $r_{j,t}$ denote the rental cost in neighborhood j and period t. The amenity in each neighborhood may be time varying, and we denote it $a_{j,t}$. Finally, we assume that there is a household- and period-specific idiosyncratic preference shock, $\varepsilon_{i,j,t}$, for household i in neighborhood j and period t.

The flow of utility for household i residing in neighborhood j at period t is $u_{j,t} + \varepsilon_{i,j,t}$, where $u_{j,t}$ depends on consumption and the amenity

$$u_{j,t} = g\left(a_{j,t}, y_t - r_{j,t}\right)$$

and where y_t is the (exogenous) income in period t.

At the beginning of each period t, the idiosyncratic preference shock is revealed. There is then another household-specific idiosyncratic draw: with probability $1 - \theta$, the household can freely relocate to any other neighborhood within the period. This relocation shock is a convenient way to capture the presence of moving rigidities. This formalization has two important properties. First, there is an exogenous and representative share θ of nonmovers in each period. Second, and in contrast with implications of the common assumption of fixed moving costs (Bayer et al. 2016), the location choice of a possible mover is not tied to their previous

location.[32] One interpretation of the option to move being a random draw is that the psychological cost of considering relocation is very high, and only an external event can force the household to pay such a cost. For instance, a liquidity shock may force the absentee landlord to sell the property—the equivalent of exogenous firing in models of labor search. Alternatively, the household may be affected by life-cycle shocks (e.g., the birth of a child).

Letting β denote the discount factor, the value of residing in neighborhood j and period t for household i is

$$U_{i,j,t} = u_{j,t} + \varepsilon_{i,j,t} + \beta\theta E_t U_{i,j,t+1} + \beta(1 - \theta)E_t V_{t+1},$$

where V_{t+1} is the value function for a household with the opportunity to relocate at $t + 1$. A household with the opportunity to relocate considers the path of future (and possibly stochastic) amenities $\{a_{j,t}\}_{j,t}$ as given and maximizes

$$V_t = \max_j\left\{u_{j,t} + \varepsilon_{i,j,t} + \beta\theta E_t U_{i,j,t+1}\right\} + \beta(1 - \theta)E_t V_{t+1}.$$

We assume, as is common in models of residential sorting (Bayer et al. 2016), that idiosyncratic preferences, $\varepsilon_{i,j,t}$, are distributed across households along a type-I extreme-value distribution, such that the fraction of households opting for neighborhood j if they have such an option in period t, $n_{j,t}^*$, follows a logit model:

$$n_{j,t}^* = \frac{e^{\sum_{\tau=0}^{\infty}(\beta\theta)^{\tau}E_t u_{j,t+\tau}}}{\sum_j e^{\sum_{\tau=0}^{\infty}(\beta\theta)^{\tau}E_t u_{j,t+\tau}}}. \tag{1}$$

Households need only consider the states of nature along which no further reoptimization is possible; these occurrences are the only ones in which the current relocation decision matters—see appendix B.1 for the derivation of equation (1) from the household optimization problem.

The equilibrium is given by (i) a sequence of neighborhood choices, resulting from the previous household optimization problem, and (ii) a sequence of rental prices, $r_{j,t}$, which adjust to equate land supply and land demand in each neighborhood and each period. We assume that households are perfectly rational. In each period, households form correct beliefs about the path of future exogenous amenities as well as current and future demand for neighborhoods, given the state of the economy. In effect, households have correct beliefs about the evolution of the endogenous amenity, $\{s_{j,t}\}$. Households are of measure zero; they take demand for neighborhoods and rental prices as given.

---

[32] The possibility to relocate is orthogonal to the current location and to the relative valuations of the different neighborhoods in the city. Conditional on being able to move, however, the household decision to stay or move will account for relative preferences for neighborhoods.

B.   Equilibrium with Two Types of Households

Consider that the city is populated by two types of households $o \in \{H, L\}$ in proportion $(\eta, 1 - \eta)$, differing only along their income, where $y^L < y^H$.[33] The demand for neighborhood j is given by equation (1),

$$
\begin{cases}
\ln\left(n_{j,t}^{H*}\right) = \sum_{\tau=0}^{\infty} (\beta\theta)^{\tau} E_t u_{j,t+\tau}^{H} - \ln\left(\sum_j e^{\sum_{\tau=0}^{\infty}(\beta\theta)^{\tau} E_t u_{j,t+\tau}^{H}}\right) \\
\ln\left(n_{j,t}^{L*}\right) = \sum_{\tau=0}^{\infty} (\beta\theta)^{\tau} E_t u_{j,t+\tau}^{L} - \ln\left(\sum_j e^{\sum_{\tau=0}^{\infty}(\beta\theta)^{\tau} E_t u_{j,t+\tau}^{L}}\right)
\end{cases},
$$

where $(n_{j,t}^{H*}, n_{j,t}^{L*})$ are the fractions of each type of household opting for neighborhood j at time t. Letting $(n_{j,t}^{H}, n_{j,t}^{L})$ denote the fractions of each type of household residing in neighborhood j at time t, we have

$$
\begin{cases}
n_{j,t}^{H} = \theta n_{j,t-1}^{H} + (1 - \theta) n_{j,t}^{H*} \\
n_{j,t}^{L} = \theta n_{j,t-1}^{L} + (1 - \theta) n_{j,t}^{L*}
\end{cases},
$$

and the land market equilibrium implies that, in each period t and neighborhood j,

$$
\eta n_{j,t}^{H} + (1 - \eta) n_{j,t}^{L} = n_j,
$$

where n_j is the fixed land supply in neighborhood j. We can use these equilibrium conditions in periods t and $t - 1$ in order to express the whole problem as a function of the share $s_{j,t}$ of type-L households among households living in neighborhood j and period t.[34] By subtracting the respective demand schedules, we obtain[35]

---

[33] We allow for this share of workers of different skill types to fluctuate over time, following migration choices, in app. B.3, and show that the relative demand for neighborhoods across types would remain unchanged. We also consider an extension with many household types and elastic land supply in app. B.3, and we show that the assumption that land supply is inelastic and constant over time is not crucial to derive the dynamic demand for neighborhoods but is required to identify the model with our data.

[34] The reader interested in the detailed derivation of the demand schedule can refer to app. B.2. Note that we normalize the total number of households to be equal to 1 without loss of generality, i.e.,

$$
\eta \sum_j n_{j,t}^{H} + (1 - \eta) \sum_j n_{j,t}^{L} = 1.
$$

[35] We define $F(x) = \ln(1 - x) - \ln(x)$ and

$$
\mu_t = \ln\left(\sum_j e^{\sum_{\tau=0}^{\infty}(\beta\theta)^{\tau} E_t u_{j,t+\tau}^{L}}\right) - \ln\left(\sum_j e^{\sum_{\tau=0}^{\infty}(\beta\theta)^{\tau} E_t u_{j,t+\tau}^{H}}\right) + \ln\left(\frac{\eta}{1-\eta}\right).
$$

The quantity μ_t captures the average relative welfare of each type in period t.

$$F\left(\frac{s_{j,t} - \theta s_{j,t-1}}{1 - \theta}\right) = \sum_{\tau=0}^{\infty}(\beta\theta)^{\tau}E_t\left(u_{j,t+\tau}^{H} - u_{j,t+\tau}^{L}\right) + \mu_t.$$

We can thus write the dynamic demand equation for neighborhood j as

$$F\left(\frac{s_{j,t} - \theta s_{j,t-1}}{1 - \theta}\right) - \beta\theta E_t F\left(\frac{s_{j,t+1} - \theta s_{j,t}}{1 - \theta}\right) = u_{j,t}^{H} - u_{j,t}^{L} + \nu_t, \qquad (2)$$

where F is a decreasing function, and $\nu_t = \mu_t - E_t\mu_{t+1}$ is a time-fluctuating variable that captures dynamics in the relative welfare of type-L residents and does not vary across neighborhoods.

Equation (2) characterizes the relative demand for neighborhood j. The right-hand side is the contemporaneous valuation of living in neighborhood j for an average type-H household relative to an average type-L household. The left-hand side is the relative demand for this neighborhood. A positive shock to the relative valuation of neighborhood j at time t, $u_{j,t}^{H} - u_{j,t}^{L}$, induces an instantaneous decrease in the fraction of low-skilled workers residing in neighborhood j. This decrease is, however, tempered by relocation rigidities: (i) some residents have not been able to respond to the change in valuation by relocating; and (ii) the residents who do have the opportunity to relocate account for the fact they may not be able to relocate in subsequent periods. With $\theta = 0$, the program of residents collapses to a static problem. They do not need to worry—in the current period—about their location in subsequent periods, and they fully adjust to any changes in their valuation of neighborhoods. With $\theta > 0$, the demand depends on the past and future expected allocation of households. The next section discusses the identification and estimation of equation (2) in the data.

V. The Persistence of Residential Sorting

In this section, we discuss the identification of the relative demand for a neighborhood in the data, present the estimates, and discuss counterfactual exercises.

A. Identification

We estimate equation (2) using the observed neighborhood composition in 142 closed cities, indexed by c, over five waves (1971, 1981, 1991, 2001, and 2011). The identification requires the following assumptions. We suppose that utility features complementarity between the neighborhood amenity, $a_{j,c,t}$, and the consumption of the numeraire and that household types differ only along their fixed income y. Specifically, we let the utility of a type-o household living in neighborhood j of city c be

$$u^o_{j,c,t} = a_{j,c,t}(y^o - r_{j,c,t}).$$

We further assume that the neighborhood amenity, $a_{j,c,t} = h(a_{j,c}, \chi_{j,c,t}, s_{j,c,t})$, is a function of (i) a constant, city-wide exogenous amenity, $a_{j,c}$; (ii) a neighborhood amenity shock, $\chi_{j,c,t}$, satisfying $E_{t-1}[\chi_{j,c,t}] = 0$ and observed by agents at the start of period t (before the relocation choice); and (iii) an endogenous amenity proxied by the current composition of the neighborhood, $s_{j,c,t}$.[36] The estimation of the relative demand for neighborhoods requires that we specify a functional form for the valuation of neighborhood j in city c. Our data do not include resident flows between neighborhoods (as in, e.g., Bayer et al. 2016), even at the aggregate level. Therefore, we can identify only a function h that is separable in its components. We show in appendix B.2 that demand for neighborhood j can then be written as

$$F\left(\frac{s_{j,c,t} - \theta s_{j,c,t-1}}{1 - \theta}\right) - \beta\theta F\left(\frac{s_{j,c,t+1} - \theta s_{j,c,t}}{1 - \theta}\right)$$
$$= \alpha_0 + \alpha_1 s^\gamma_{j,c,t} + b_{j,c} + \zeta_{j,c,t} + \nu_{c,t}, \tag{S3}$$

where $b_{j,c}$ is a neighborhood fixed effect, unobserved to the econometrician, and the noise $\zeta_{j,c,t}$ is a linear combination of the neighborhood amenity shock and a prediction error, $F((s_{j,c,t+1} - \theta s_{j,c,t})/(1 - \theta)) - E_t F((s_{j,c,t+1} - \theta s_{j,c,t})/(1 - \theta))$, thus satisfying $E_{t-1}[\zeta_{j,c,t}] = 0$.

In specification (S3), the parameters of interest are θ, β, and the vector of parameters characterizing the relative valuation of neighborhoods, (α, γ).[37] The estimation needs to absorb city/wave fixed effects, $\nu_{c,t}$, and neighborhood fixed effects, $b_{j,c}$.

We estimate specification (S3) by generalized method of moments. The (local) identification of the discount factor, β, the relocation rigidities, θ, and the preferences for neighborhood quality are formally derived in appendix B.5. Intuitively, the identification of preferences for neighborhood quality relies on variation in the share of low-skilled workers across neighborhoods at a point in time, $s_{j,c,t}$. By contrast, the identification of the discount factor and relocation rigidities comes from the dynamics of residential sorting. The change in residential sorting from one period to another is important in pinning down both parameters. The acceleration/slowdown in residential sorting (the change in the change in sorting) is

---

[36] The composition of the neighborhood is used as a proxy for the wide range of neighborhood effects that may affect the perceived "neighborhood quality." Durlauf (2004) and Rosenthal and Ross (2015) provide excellent overviews of neighborhood effects affecting residential choices. These effects may include school quality (Durlauf 1996), quality of the housing stock (Rosenthal 2008), or preferences to live among workers of similar or higher income groups (Guerrieri, Hartley, and Hurst 2013) or the same ethnic group (Card, Mas, and Rothstein 2008).

[37] We study the dynamic properties of the system of equations (spec. [S3]) in app. B.4.

crucial in separating the specific role of relocation rigidities. The following thought experiment helps understand this last argument. Consider one neighborhood that is expected to be polluted until $t + 1$ and clean in $t + 2$. High relocation rigidities will induce very similar relocation patterns in t, $t + 1$, or $t + 2$: agents anticipate that they may not be able to move in the future and relocate as soon as they are given the opportunity. With low relocation rigidities, agents will instead mostly relocate in $t + 2$. High relocation rigidities are thus identified through sluggish dynamics.

A concern with the estimation of specification (S3) is that, in the data, the share of low-skilled workers, $s_{j,c,t}$, may be correlated with the unobserved noise $\zeta_{j,c,t}$. Neighborhood dynamics correlate with the neighborhood amenity shock, unobserved to the econometrician. We thus exclude the shares of low-skilled workers, $s_{j,c,t}$, from the set of instruments, and we create a set of new instruments exploiting the temporary disamenity induced by historical pollution. Pollution had an impact on the initial stock of low-skilled workers across neighborhoods. Given the role of neighborhood dynamics in identifying equation (S3), we interact past pollution, as a shifter for the initial share of low-skilled workers, with wave fixed effects. The set of instruments allows us to isolate variations in the share of low-skilled workers in 1971, in its growth from one period to the other, and in its acceleration/slowdown (as shown in figs. 10, 11).

In parallel, we control for the interactions of (i) the initial share of social housing in 1971 and (ii) bombing intensity during the German Blitz, with wave fixed effects, in order to clean for dynamics induced by social housing policies (e.g., the Housing Act 1980) and urban renewal.[38] We also control flexibly for nonlinear dynamics linked to the presence of amenities in a neighborhood by interacting a measure of predicted amenities (using all controls of table 2, col. 6, to predict the share of low-skilled workers in 1971) with wave fixed effects. The identifying assumption is that historical pollution is orthogonal to neighborhood amenity shocks and affects only relative demand for neighborhoods through inherited neighborhood composition, conditional on time-varying controls for social housing, past exposure to bombings, and predicted neighborhood amenities. A premise is that industrial pollution from coal burning disappears from cities after 1971.

B. Structural Estimates

The estimates for specification (S3) are reported in table 7. Column 1 reports the baseline specification where neighborhood preferences are linear in neighborhood composition. The share of low-skilled workers

---

[38] We study the respective roles of social housing and bombing intensity on neighborhood sorting in apps. G and H.

TABLE 7
Relative Demand for Neighborhoods—Structural
Estimation for Specification (S3)

	(1)	(2)	(3)	(4)	(5)
Relocation rigidity θ	.4782	.4777	.4706	.4730	.4897
	(.0229)	(.0221)	(.0314)	(.0209)	(.0225)
Discount factor β	.9303	.9000	.6312	.8171	1.365
	(.3783)	(.3152)	(.9056)		(.2660)
Preferences α_1	−2.869	−2.893	−3.652	−3.122	−5.285
	(1.051)	(.9147)	(2.062)	(.7225)	(1.590)
Preferences γ	1	1	1	1	1.170
					(.3079)
Observations	16,284	16,284	16,284	16,284	16,284
Instruments	(I_1)	(I_2)	(I_3)	(I_1)	(I_2)

Note.—Standard errors are in parentheses and clustered at the parish × wave level. The unit of observation is a lower layer super output area in 1981, 1991, or 2001. The estimation is performed using a one-step generalized method of moments estimator, with all endogenous variables included as instruments, except neighborhood composition variables, which are replaced by measures of historical pollution × wave fixed effects. The set of instruments I_1 include our baseline measure of historical pollution interacted with wave fixed effects. The set of instruments I_2 include pollution deciles interacted with wave fixed effects. The set of instruments I_3 include distance to waterways interacted with wave fixed effects. All specifications include city/wave and neighborhood fixed effects and the interactions of (i) bombing intensity during the German Blitz and (ii) the initial share of social housing in 1971 with time fixed effects, and (iii) a measure of predicted amenities (using all controls in table 2, col. 6, to predict the share of low-skilled workers in 1971). The relocation rigidity corresponds to a yearly probability to be able to relocate of 7%; the 10-year discount rate β corresponds to a yearly discount rate of 0.993 (col. 1).

lowers the valuation of the neighborhood, which generates a multiplier effect. One additional standard deviation in the share of low-skilled workers (about 7 percentage points) reduces the relative valuation of a neighborhood by about 0.20, which would trigger—off equilibrium and at the steady state—a further decrease in the share of high-skilled workers of 3 percentage points. This finding is illustrated in figure A8. The persistence in neighborhood dynamics derives from the existence of neighborhood effects, combined with non-negligible moving rigidities. Past neighborhood composition directly influences current neighborhood composition, as many residents will not be given the opportunity to relocate. Movers anticipate this long-lasting effect on neighborhood composition and adjust their location choices accordingly. We quantify this indirect, forward-looking effect in the next section.

The probability of being given the opportunity to relocate over a period of 10 years, $1 - \theta$, is estimated to be 52%. Assuming that residents always move when given the possibility (they would almost certainly do so in a city with many neighborhoods), this estimate implies an average housing tenure of about 14 years, consistent with average rates of turnover observed in the housing market. The annualized discount rate, estimated

to be around 0.7%, is in the low range of risk-free rates observed over the past 50 years (in spite of a recent drop in such rates; Del Negro et al. 2019).

We then run a series of robustness checks. We add the set of dummies for different pollution deciles interacted with time trends as instruments, in order to exploit the nonlinearities in neighborhood dynamics documented in figure 11 (see table 7, col. 2). We replace the pollution-based instruments by distance to waterways interacted with wave fixed effects in order to replicate the intuition behind our IV strategy in column 3. We calibrate the discount factor as induced by a 2% annualized discount rate and estimate the remaining parameters (col. 4). Finally, we estimate a nonlinear specification in column 5. The baseline specification is robust to the use of a different set of instruments; the identification of the relocation rigidity θ does not appear to be tied to the identification of the discount factor; the parameter γ is not significantly different from 1. For these reasons, we consider the estimates presented in column 1 as the baseline estimated parameters in the following simulations and counterfactual experiments.

C. Dynamics of Neighborhood Segregation

To consider the in-sample performance of the model, we simulate neighborhood dynamics between 1991 and 2011 using the structural estimates reported in column 1 of table 7 and the actual shares of low-skilled workers in 1971 and 1981. Figure 12 shows the model performance in explaining the transitional dynamics between 1971 and 2011. There is a secular and exogenous decrease in the share of low-skilled workers. This decrease is, however, more pronounced for neighborhoods with low historical pollution exposure, where neighborhood composition does not anchor expectations about future amenities. In effect, the relationship between past pollution and deprivation is at least as strong in 2011 as in 1971, when industrial pollution abruptly waned from urban centers. We report the observed shares of low-skilled workers in 2011 to provide some visual evidence of the model fit: the model performs well in reproducing the general neighborhood dynamics.

Next, we introduce two counterfactual experiments to help us understand the respective role of preferences and frictions in the dynamics of segregation. In a first counterfactual experiment, we simulate neighborhood dynamics in a model where preferences are set to be orthogonal to current neighborhood composition. More specifically, we set the parameter α_1 to zero and adjust the neighborhood fixed amenity to keep the right-hand side of specification (S3) unchanged in 1971. In figure A9a, we report the simulated share of low-skilled workers in 2011. The secular decrease in the share of low-skilled workers between 1971 and 2011 is
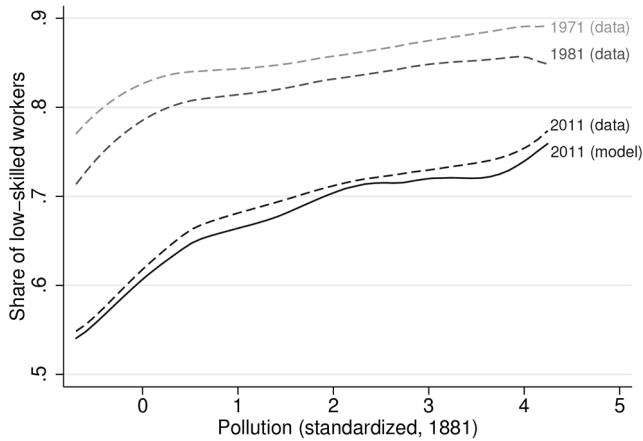
FIG. 12.—Simulated dynamics (1971–2011) using the structural estimates. Shown are the average shares of low-skilled workers as a function of past pollution, using a local polynomial smoothing. We report neighborhood dynamics, as estimated using the structural estimates reported in column 1 of table 6 and the observed shares of low-skilled workers in 1971 and 1981. The solid line represents simulated shares in 2011, while dashed lines represent observed shares of low-skilled workers in 1971, 1981, and 2011 (shown from lighter to darker gray, respectively).

much larger in formerly highly polluted neighborhoods: neutralizing neighborhood effects is sufficient to turn around the dynamics of segregation, with the remaining correlation between past pollution and deprivation being entirely explained by fixed geographic amenities. In a second counterfactual experiment, we keep preferences as in the baseline model and instead modify the extent of rigidities in the relocation process. More precisely, we increase the annual probability to be given the opportunity to relocate from 7% to 13% (corresponding to $\theta = 0.25$), such that most residents have the opportunity to relocate between two census waves. We report the simulated dynamics in figure A9b. Again, the simulated dynamics sharply differ from the baseline scenario with a quick reversion to the mean in formerly highly polluted neighborhoods. These simulations shed light on the crucial role of the combination of neighborhood effects and relocation rigidities in the dynamics of segregation.

D. *Quantifying the Impact of Past Pollution and Social Housing*

The Clean Air Acts of 1952 and 1968 penalized the emissions of grit, dust, and dark smoke in urban centers and succeeded in reducing emissions from burning coal. However, past pollution exposure leaves a long
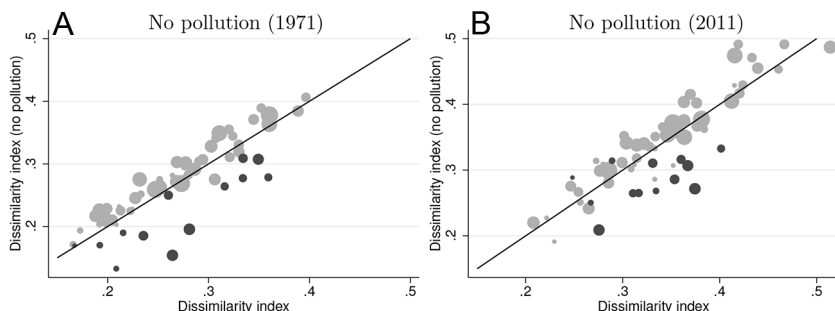
FIG. 13.—Neighborhood segregation across cities in the absence of pollution in 1971 (*A*) and 2011 (*B*). Compared are measures of segregation across cities in two experiments: (i) simulated dynamics based on actual neighborhood compositions in 1971 and 1981 (*x*-axis) and simulated dynamics based on counterfactual neighborhood compositions in 1971 and 1981 (no pollution; *y*-axis). We construct the counterfactual neighborhood compositions by subtracting the effect of pollution, as measured in table 4 (cols. 1, 2). Each city is represented by a circle whose size is proportional to its number of lower layer super output areas (LSOAs) within the sample and whose color shows the importance of pollution at the city level. Dark gray circles represent cities in the top quartile of pollution; light gray circles represent cities in the lower quartiles. The measure of segregation used is one of evenness, that is, dissimilarity: $(1/2J_c)\Sigma_{j=1}^{J}|(s_{j,c}/s_c) - (1 - s_{j,c}/1 - s_c)|$, where $s_{j,c}$ is the share of low-skilled workers in LSOA j of city c, s_c is the average share of low-skilled workers in city c, and J_c is the number of LSOAs in city c.

shadow, arising from the inherited neighborhood composition and the subsequent persistence of neighborhood sorting.

We now quantify the magnitude of this effect, among all other drivers of residential segregation (e.g., fixed neighborhood amenities). We first construct a counterfactual distribution of residents in 1971 and 1981 across neighborhoods by subtracting the effect of pollution, as measured in table 4 (cols. 1, 2). We then compare the simulated dynamics based on such counterfactual neighborhood composition to the simulated dynamics based on actual neighborhood composition in 1971 and 1981 (see fig. 13). To this end, we construct a measure of evenness, dissimilarity, for each city in our sample.[39] Figure 13 shows the differences between the measures of segregation in the two experiments. Dark gray circles represent cities in the top quartile of pollution; light gray circles represent cities in the lower quartiles of pollution. In 1971, just after the Clean Air Act of 1968, the dissimilarity measure is about 0.07 (about 1 SD) lower for heavily polluted cities when we shut down the effect of pollution. In 2011, the effect of pollution on the dissimilarity measure remains equally large for heavily polluted cities.

---

[39] The dissimilarity measure is defined as $(1/2J_c)\Sigma_{j=1}^{J}|(s_{j,c}/s_c) - (1 - s_{j,c}/1 - s_c)|$, where $s_{j,c}$ is the share of low-skilled workers in LSOA j of city c, s_c is the average share of low-skilled workers in city c, and J_c is the number of LSOAs in city c.

Social housing policies, such as the Housing Act 1980, could have had a role in tempering or fostering neighborhood segregation between 1971 and 2011. We quantify their impact by considering an experiment in which the share of social housing in 1971 imposes a minimum level for the fraction of low-skilled workers at the LSOA level. We simulate the dynamics between 1991 and 2011 in this scenario, and we report the correlation between past pollution and the simulated deprivation in figure 14A. We also build on our previous analysis and compare the dissimilarity measures across cities between the baseline scenario and the social housing scenario in 2011 (see fig. 14B).

The locational rigidities induced by a fixed stock of social housing at the LSOA level reduce neighborhood segregation in 2011, compared to the baseline. The dissimilarity measure is about 1 SD lower, with locational rigidities for those cities with an above-median stock of social housing in 1971. The reason for this counterintuitive finding is that social housing is negatively correlated with past pollution in 1971 and aligns only with deprivation from 1991 onward (see app. G). A policy that would confine social housing to its predetermined location would thus limit the extent to which well-connected, attractive locations would further gentrify (as documented, e.g., in Guerrieri, Hartley, and Hurst 2013; Baum-Snow and Hartley 2020; and Couture and Handbury 2020).
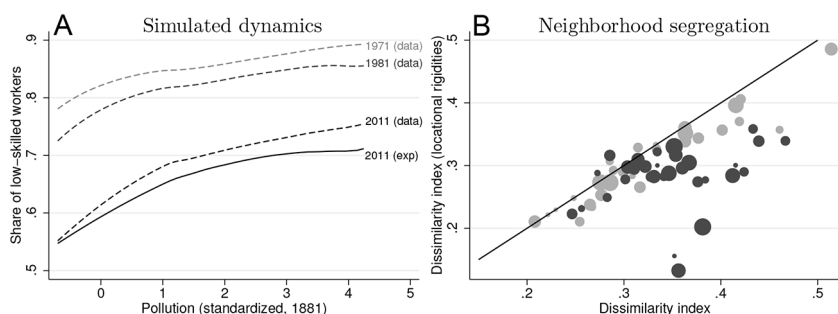

Fig. 14.—Neighborhood dynamics and segregation with fixed social housing. A, The average shares of low-skilled workers as a function of past pollution when there are some locational rigidities—as induced by imposing a constant, lower threshold for the share of low-skilled workers at the lower layer super output area (LSOA) level (corresponding to the minimum between this share and fractions of low-skilled workers in 1971 and 1981). B, Comparison of measures of segregation across cities in two experiments: (i) unrestricted simulated dynamics and (ii) simulated dynamics with locational rigidities (y-axis). Each city is represented by a circle whose size is proportional to its number of LSOAs within the sample and whose color shows the importance of social housing at the city level. Dark gray circles represent cities with above-median shares of social housing in 1971; light gray circles represent cities with below-median shares of social housing in 1971. The measure of segregation used is one of evenness, that is, dissimilarity: $(1/2J_c)\Sigma_{j=1}^{J_c}|(s_{j,c}/s_c) - (1 - s_{j,c}/1 - s_c)|$, where $s_{j,c}$ is the share of low-skilled workers in LSOA j of city c, s_c is the average share of low-skilled workers in city c, and J_c is the number of LSOAs in city c.

VI. Conclusion

This paper presents a plausible explanation for what was, until now, an anecdotal observation that the east sides of formerly industrial cities in the Western Hemisphere tend to be poorer than the west sides. With rising coal use in the heyday of industrialization, pollution became a major environmental disamenity in cities. An unequal distribution of pollution exposure induced a sorting process that left lower classes in polluted neighborhoods. Our empirical analysis relies on precise pollution estimates and identifies neighborhood sorting at a highly local level: the east/west gradient reflects a drift in pollution at the city level, but the relationship between atmospheric pollution and neighborhood composition materializes at a much more local level.

We first use data from the time before coal became the major energy technology in 1817 as well as data around the peak time of coal use in 1881 to show that rising pollution set off the process of residential sorting. Next, we look at the long-run consequences of this initial sorting and find that neighborhood segregation is surprisingly persistent. Finding these highly persistent effects is remarkable since industrial pollution slowed down during the twentieth century and mostly stopped in the late 1960s with the introduction of a second, stricter Clean Air Act. There exists no correlation between past industrial pollution and the relatively mild contemporary pollution in England, suggesting that other forces have sustained neighborhood segregation over time. We use a quantitative model with relocation rigidities and neighborhood effects to estimate a dynamic demand equation for neighborhoods. Our structural estimates imply nonlinear transitional dynamics that relate to the literature on tipping dynamics (Card, Mas, and Rothstein 2008).

Our findings hold at least two important implications. First, the success of urban policies to revitalize deprived areas may depend on the initial level of deprivation. As suggested by our findings, very deprived neighborhoods may need a larger push to attract richer residents. This observation leads to a second implication for countries such as China where pollution currently presents a major challenge. Besides the well-documented short-run effects of pollution exposure on health, there are long-run consequences of uneven pollution exposure across space: pollution induces spatial inequalities that far outlive deindustrialization.

References

Anderson, Michael L. 2020. "As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality." *J. European Econ. Assoc.* 18 (4): 1886–927.

Banzhaf, H. Spencer, and Randall P. Walsh. 2008. "Do People Vote with Their Feet? An Empirical Test of Tiebout." *A.E.R.* 98 (3): 843–63.

Barreca, Alan, Karen Clay, and Joel Tarr. 2014. "Coal, Smoke, and Death: Bituminous Coal and American Home Heating." Working Paper no. 20648, NBER, Cambridge, MA.

Baum-Snow, Nathaniel, and Daniel A. Hartley. 2020. "Accounting for Central Neighborhood Change, 1980–2010." *J. Urban Econ.* 117:103228.

Bayer, Patrick, Robert McMillan, Alvin Murphy, and Christopher Timmins. 2016. "A Dynamic Model of Demand for Houses and Neighborhoods." *Econometrica* 84 (3): 893–942.

Beach, Brian, and W. Walker Hanlon. 2018. "Coal Smoke and Mortality in an Early Industrial Economy." *Econ. J.* 128:2652–75.

Bleakley, Hoyt, and Jeffrey Lin. 2012. "Portage and Path Dependence." *Q.J.E.* 127 (2): 587–644.

Brimblecombe, Peter. 2006. "The Clean Air Act after 50 Years." *Weather* 61 (11): 311–14.

Brooks, Leah, and Byron Lutz. 2019. "Vestiges of Transit: Urban Persistence at a Microscale." *Rev. Econ. and Statis.* 101 (3): 385–99.

Brueckner, Jan K., and Stuart S. Rosenthal. 2009. "Gentrification and Neighborhood Housing Cycles: Will America's Future Downtowns Be Rich?" *Rev. Econ. and Statis.* 91 (4): 725–43.

Card, David, Alexandre Mas, and Jesse Rothstein. 2008. "Tipping and the Dynamics of Segregation." *Q.J.E.* 123 (1): 177–218.

Chay, Kenneth Y., and Michael Greenstone. 2005. "Does Air Quality Matter? Evidence from the Housing Market." *J.P.E.* 113 (2): 376–424.

Chen, Shuai, Paulina Oliva, and Peng Zhang. 2017. "The Effect of Air Pollution on Migration: Evidence from China." Working Paper no. 24036, NBER, Cambridge, MA.

Clark, G. Kitson. 1962. *The Making of Victorian England.* Cambridge, MA: Harvard Univ. Press.

Clay, Karen, Joshua Lewis, and Edson Severnini. 2016. "Canary in a Coal Mine: Infant Mortality, Property Values, and Tradeoffs Associated with Mid-20th Century Air Pollution." Working Paper no. 22155, NBER, Cambridge, MA.

Conley, Timothy G. 1999. "GMM Estimation with Cross Sectional Dependence." *J. Econometrics* 92 (1): 1–45.

Cook, Laurence M. 2003. "The Rise and Fall of the *Carbonaria* Form of the Peppered Moth." *Q. Rev. Biology* 78 (4): 399–417.

———. 2018. "Records of Industrial Melanism in British Moths." *Biological J. Linnean Soc.* 125:862–66.

Couture, Victor, and Jessie Handbury. 2020. "Urban Revival in America." *J. Urban Econ.* 119:103267, https://doi.org/10.1016/j.jue.2020.103267.

Crafts, Nicholas. 2014. "Economic Growth during the Long Twentieth Century." In *The Cambridge Economic History of Modern Britain. Volume II: 1870 to the Present*, edited by R. Floud, J. Humphries, and P. Johnson, 26–59. Cambridge, MA: Cambridge Univ. Press.

Crafts, Nicholas, and C. Knick Harley. 1992. "Output Growth and the British Industrial Revolution: A Restatement of the Crafts-Harley View." *Econ. Hist. Rev.* 45 (4): 703–30.

Davis, Donald R., and David E. Weinstein. 2002. "Bones, Bombs, and Break Points: The Geography of Economic Activity." *A.E.R.* 92 (5): 1269–89.

Del Negro, Marco, Domenico Giannone, Marc P. Giannoni, and Andrea Tambalotti. 2019. "Global Trends in Interest Rates." *J. Internat. Econ.* 118:248–62.

Depro, Brooks, Christopher Timmins, and Maggie O'Neil. 2015. "White Flight and Coming to the Nuisance: Can Residential Mobility Explain Environmental Injustice?" *J. Assoc. Environmental and Res. Econ.* 2 (3): 439–68.

Deryugina, Tatyana, Garth Heutel, Nolan H. Miller, David Molitor, and Julian Reif. 2019. "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction." *A.E.R.* 109 (12): 4178–219.

Durlauf, Steven N. 1996. "A Theory of Persistent Income Inequality." *J. Econ. Growth* 1:75–93.

———. 2004. "Neighborhood Effects." In *Handbook of Regional and Urban Economics*, vol. 4, edited by J. Vernon Henderson and J. F. Thisse, 2173–242. Amsterdam: Elsevier.

Freeman, Richard, Wenquan Liang, Ran Song, and Christopher Timmins. 2019. "Willingness to Pay for Clean Air in China." *J. Environmental Econ. and Management* 94:188–216.

Graff Zivin, Joshua, and Matthew Neidell. 2012. "The Impact of Pollution on Worker Productivity." *A.E.R.* 102 (7): 3652–73.

———. 2013. "Environment, Health, and Human Capital." *J. Econ. Literature* 51 (3): 689–730.

Guerrieri, Veronica, Daniel Hartley, and Erik Hurst. 2013. "Endogenous Gentrification and Housing Price Dynamics." *J. Public Econ.* 100:45–60.

Hanlon, W. Walker. 2019. "Coal Smoke and the Costs of the Industrial Revolution." *Econ. J.* 130:462–88.

Heblich, Stephan, Stephen J. Redding, and Daniel M. Sturm. 2020. "The Making of the Modern Metropolis: Evidence from London." *Q.J.E.* 135 (4): 2059–133.

Herrnstadt, Evan, Anthony Heyes, Erich Muehlegger, and Soodeh Saberian. 2019. "Air Pollution as a Cause of Violent Crime: Evidence from Los Angeles and Chicago." Working paper, Univ. Ottawa

Hornbeck, Richard, and Daniel Keniston. 2017. "Creative Destruction: Barriers to Urban Growth and the Great Boston Fire of 1872." *A.E.R.* 107:1365–98.

Kanefsky, John, and John Robey. 1980. "Steam Engines in 18th-Century Britain: A Quantitative Assessment." *Tech. and Culture* 21 (2): 161–86.

Kellett, John R. 1969. *The Impact of Railways on Victorian Cities.* London: Routledge & Kegan Paul.

Kuminoff, Nicolai V., V. Kerry Smith, and Christopher Timmins. 2013. "The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets." *J. Econ. Literature* 51 (4): 1007–62.

Lamb, Hubert H. 1972. *British Isles Weather Types and a Register of Daily Sequence of Circulation Patterns 1861–1971.* Vol. 116 of the Meteorological Office Geophysical Memoirs. London: Her Majesty's Stationery Office.

Lavy, Victor, Avraham Ebenstein, and Sefi Roth. 2016. "The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution." *American Econ. J. Appl. Econ.* 8 (4): 36–65.

Lawton, Richard. 1972. "An Age of Great Cities." *Town Planning Rev.* 43 (3): 199–224.

Lee, Sanghoon, and Jeffrey Lin. 2018. "Natural Amenities, Neighborhood Dynamics, and Persistence in the Spatial Distribution of Income." *Rev. Econ. Studies* 85:663–94.

Logan, Trevon, and John M. Parman. 2017. "The National Rise in Residential Segregation." *J. Econ. Hist.* 77 (1): 127–170.

Maw, Peter, Terry Wyke, and Alan Kidd. 2012. "Canals, Rivers, and the Industrial City: Manchester's Industrial Waterfront, 1790–1850." *Econ. Hist. Rev.* 65 (4): 1495–523.

Mosley, Stephen. 2013. *The Chimney of the World: A History of Smoke Pollution in Victorian and Edwardian Manchester.* London: Routledge.

Murphy, Kevin M., and Robert H. Topel. 2002. "Estimation and Inference in Two-Step Econometric Models." *J. Bus. and Econ. Statis.* 20 (1): 88–97.

Musson, A. E. 1976. "Industrial Motive Power in the United Kingdom, 1800–70." *Econ. Hist. Rev.* 29 (3): 415–39.

Nuvolari, Alessandro, Bart Verspagen, and Nick von Tunzelmann. 2011. "The Early Diffusion of the Steam Engine in Britain, 1700–1800: A Reappraisal." *Cliometrica* 5 (3): 291–321.

Ravenstein, Ernest G. 1885. "The Laws of Migration." *J. Statis. Soc. London* 48:167–235.

Redding, Stephen J., and Daniel M. Sturm. 2016. "Estimating Neighborhood Effects: Evidence from War-Time Destruction in London." Working paper, Dept. Econ., Princeton Univ.

Rosenthal, Stuart S. 2008. "Old Homes, Externalities, and Poor Neighborhoods. A Model of Urban Decline and Renewal." *J. Urban Econ.* 63:816–40.

Rosenthal, Stuart S., and Stephen Ross. 2015. "Change and Persistence in the Economic Status of Neighborhoods and Cities." In *Handbook of Regional and Urban Economics*, vol. 5, edited by J. Vernon Henderson, Gilles Duranton, and William C. Strange, 1048–120. Amsterdam: Elsevier.

Schelling, Thomas C. 1971. "Dynamic Models of Segregation." *J. Math. Soc.* 1 (2): 143–86.

Shaw-Taylor, Leigh, and E. Anthony Wrigley. 2014. "Occupational Structural and Population Change." In *Cambridge Economic History of Modern Britain. Volume 1: Industrialisation 1700–1860*, edited by Roderick Floud, Jane Humphries, and Paul Johnson, 53–88. Cambridge, MA: Cambridge Univ. Press.

Siodla, James. 2015. "Razing San Francisco: The 1906 Disaster as a Natural Experiment in Urban Redevelopment." *J. Urban Econ.* 89:48–61.

Villarreal, Carlos. 2014. "Where the Other Half Lives: Evidence on the Origin and Persistence of Poor Neighborhoods from New York City 1830–2012." Working paper, Dept. Econ., Univ. Chicago.

Warde, Paul. 2007. *Energy Consumption in England and Wales, 1560–2004*. Naples: Consiglio Nazionale della Ricerche.

Williamson, Jeffrey G. 1990. *Coping with City Growth during the British Industrial Revolution*. Cambridge, MA: Cambridge Univ. Press.

Wrigley, E. Anthony. 2010. "The PST System of Classifying Occupations." Working paper, Cambridge Univ.