Heblich, S. H., Trew, A., & Zylberberg, Y. (Accepted/In press). East Side Story: Historical Pollution and Persistent Neighborhood Sorting. *Journal of Political Economy*.

Peer reviewed version

East Side Story: Historical Pollution and Persistent Neighborhood Sorting*

Stephan Heblich Alex Trew Yanos Zylberberg

July 6, 2020

Abstract

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 eastwards. Past pollution explains up to 20% of the 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 in underlying this persistence.

Keywords: Neighborhood Sorting, Historical Pollution, Persistence.
JEL codes: R23, Q53, N00.

*Heblich: University of Toronto, CEP, CESifo, IfW Kiel, IZA; stephan.heblich@bristol.ac.uk; Trew: University of Glasgow; alex.trew@glasgow.ac.uk; Zylberberg: University of Bristol, CESifo, the Alan Turing Institute; yanos.zylberberg@bristol.ac.uk. We thank the editor and six referees for comments that greatly improved this paper, as well as Philipp Ager, Spencer Banzhaf, Nate Baum-Snow, Kristian Behrens, Davide Cantoni, Peter Egger, Klaus Desmet, Christian Dippel, Gilles Duranton, Oliver Falck, Ed Glaeser, Walker Hanlon, Hans Hvide, Hans Koster, Robert McMillan, Jos van Ommeren, Giacomo Ponzetto, Diego Puga, Joshua Lewis, Jeffrey Lin, Petra Moser, Steve Redding, Jorge De la Roca, Stuart Rosenthal, Max Satchell, Hannes Schwandt, William Strange, Leigh Shaw-Taylor, Jon Temple, Chris Timmins, Nico Voigtländer, and Ludgar Wössmann for useful discussions and comments. We also thank seminar participants in Bayreuth, Bergen, Bern, Boulder, Bristol, Cambridge, CERGE-EI, CUHK, Dublin, Durham, Edinburgh, Essex, HSE St. Petersburg, Kiel, Munich, NYU Stern, Odense, Passau, the Philadelphia FED, SERC, St Andrews, Stirling, Toronto, UCLA Anderson, USD, Warwick and participants at the 2016 IEB Workshop on Urban Economics, the 2016 EUEA, the 2016 UEA, the 2017 ASSA meeting and the 2017 NBER EEE meeting for helpful comments. We are grateful to a number of excellent research assistants. This work was part-funded by the ESRC through the Applied Quantitative Methods Network: Phase II, grant number ES/K006460/1. The usual disclaimer applies.

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 which affected cities during the Industrial Revolution. Pollution from historical factories accounts 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 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 only started 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 19th 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 model to construct pollution maps from the chimney locations.[1] 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 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 percentile 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 a substantial persistence in the effect of historical pollution on within-city distribution of low-skilled workers even after the

[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).

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 non-linearities, 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 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 et al., 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 OLS specification.

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 Bayer et al., 2016, for instance). This setup yields a simple dynamic equation which characterizes the relative demand for neighborhoods. A key advantage is the comparatively low data requirement; instead of individual data or flows between neighborhoods, identification only requires the aggregate share of low-skilled households in each neighborhood over three periods. Using the shares of low-skilled households from 1971 to 2011, we estimate the relative demand for neigh-

[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.

borhoods using 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 not only tied to relocation frictions, but 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 that move in each period. An additional effect arises from the forward-looking behavior of movers: they anticipate the slow future adjustment of neighborhood composition in0 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 one standard deviation, as evaluated in 1971 but also in 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 Lee and Lin (2018) who look at exogenous natural amenities as driver of neighborhood sorting. Our paper, in contrast, studies the consequences of a temporary disamenity. To the best of our knowledge, we are the first to show that the large, temporary pollution from industrial coal use modified the spatial organization of cities in the long run.[3] Related papers that look at pollution-induced sorting in a purely modern setting include Banzhaf and Walsh (2008) and Chay and Greenstone (2005).[4] Kuminoff et al. (2013) provide a broader review of the residential sorting literature. Our argument further relates to Depro et al. (2015) who argue that neighborhood sorting, rather than environmental injustice, is the reason why poor households are more exposed to environmental disamenities. Finally, our paper shares a common theme with Hanlon (2019), who argues that coal-based pollution was a significant disamenity with a strong negative impact on city size in England during the industrialization, and with Chen et al. (2017) and Freeman et al. (2019)

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

[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 et al., 2016), violent crime (Herrnstadt et al., 2019), and health (Graff Zivin and Neidell, 2013; Anderson, 2019; Deryugina et al., 2019) which relates to our research. Closely related historical assessments of the effect of coal use on health include (Barreca et al., 2014; Clay et al., 2016; Beach and Hanlon, 2018).

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 U.S. cities (Brueckner and Rosenthal, 2009; Guerrieri et al., 2013; Baum-Snow and Hartley, 2020; Couture and Handbury, 2019) which, in turn, builds on previous research on the dynamics of segregation and tipping points (Schelling, 1971; Card et al., 2008; Logan and Parman, 2017). In our context, we mostly identify a social component behind segregation (in contrast to the literature on the U.S. that 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 non-linearities 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 to digitize historical maps to create a new dataset on pollution and the structure of cities. Related to this approach is work by Hornbeck and Keniston (2017) and Siodla (2015), 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 Second World War destruction in London.

The remainder of the paper is organized as follows. Section 1 provides elements of context and describes data sources. The reduced-form evidence on neighborhood sorting and its persistence is discussed in Section 2. Section 3 develops a dynamic model of residential choice. The identification and estimation of the demand for neighborhoods are discussed in Section 4. Finally, Section 5 concludes.

1 Historical background and data

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

1.1 Historical background

The start of the Classical Industrial Revolution is dated to around 1760 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] The left panel of Figure 1 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 world wide (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]

The heavy reliance on coal between 1850 and 1960 generated unprecedented concentrations of sulphur dioxide 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

[5] As Musson (1976) shows, power derived from water wheels remained important to early nineteenth century industry—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 de-industrialization, which occurred most rapidly in the 1980s when state-owned industries were privatized.

[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 in Barreca et al. (2014), Clay et al. (2016), and Beach and Hanlon (2018).

among Manchester's working classes might at least escape the worst of the smoke." 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 induced 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 countryside flocking into the emerging industrial cities (Shaw-Taylor and Wrigley, 2014). As shown in the right panel of Figure 1, 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, located within walking distance to the new factories. The lack of suitable housing along with limited supply of clean water and sanitation created unhealthy urban slums plagued by diseases like cholera and typhoid, leading to notoriously low life expectancy (Clark, 1962). Only in the second half of the 19th 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 in Heblich et al. (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 work places (Kellett, 1969; Lawton, 1972).

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

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.

1.2 Data sources and construction

This section provides a brief summary of the different data sources; a comprehensive description can be found in Appendix C.

OS maps and geo-location of pollution sources Data on city structure and pollution sources are drawn from the 25 inch to the mile Ordnance Survey (OS) maps. 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 and 66% in 2011. The maps contain details on roads, railway, rivers, canals, 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 geo-reference 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 the chimney location.[10] Aggregate measures of coal use per worker are calculated for the following eleven industrial categories: *Brick factories*, *Foundries*, *Chemical factories*, *Mining*, *Breweries*, *Tanneries*, *Food processing*, *Textile production*, *Paper production*, *Shipbuilding* and *Wood processing* (Hanlon,

[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. We benefited from the excellent research assistance of Andreas Arbin, Nicholas Cheras, Tim Ciesla, Joshua Croghan, Qingli Fang, Joanna Kalemba, Aishwarya Kakatkar, Matthew Litherland, Filip Nemecek, Ondrej Ptacek and Sava Zgurov.

[10]The industrial category—measured with error—is the only information that we use in order to proxy for the average pollution emission associated to 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.

2019). A simple textual analysis based on 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 follows:

$$E_i = \frac{C_i \times L_i}{Ch_i}$$

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]

Pollution dispersion The previous exercise produces a map of industrial pollution sources. The *ADMS 5* dispersion model then permits us to construct a city-wide map of exposure to air pollutants (SO_2, measured in $\mu g/m^3$).[12] *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 19th 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/south-west; 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 on current chimneys which are tall, wide and have high exit velocity. By contrast, chimneys in the Industrial Revolution were between 10 and 50 meters tall with the majority being shorter than 25 meters. Moreover, the exit

[11] We report the estimated pollution emission per industry, E_i, in Appendix Table A1. 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*).

[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 Appendix Figure 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.

velocity and temperature were also lower than today. To incorporate these characteristics, we set chimney height to 25 meters in the baseline and assume an exit velocity of 4 m/s and an exit temperature of 120 degrees Celsius. 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 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). Appendix Figure A2 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.

Measure of neighborhood composition Measures of neighborhood sorting in the nineteenth century are extracted from individual records of the 1881 census. These records hold information on the structure of households and each occupant's gender, age, occupation and place of birth. There are two indicators of household

[14]In our sample of industrialized cities, the relative contribution of domestic emissions (versus industrial emissions) in explaining the distribution of pollutant concentration within 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.

[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). Appendix 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 $15\mu g/m^3$). About 2% of LSOAs—mostly in Manchester, Oldham and Liverpool—have indices of pollution above the peaks recorded in contemporary Beijing ($40\mu g/m^3$).

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 Super Output Area, LSOA), based on the organization of data collection.

Individual surveyors were given blocks to survey and they each filled in enumerator books while visiting their allocated neighborhoods: this induces is 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 geo-references of well-matched 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 pre-period, 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 micro-census 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 3-digit occupational information in recent censuses and rely on a similar classification

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

[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.

for 1817 and 1881 (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, i.e., the share of low-skilled workers among the working population. For 1817 and 1881, we collapse the 500 occupational sub-categories into 10 categories. We restrict the sample to individuals with the lowest possible measurement error, i.e., males between 25 and 55.[18] Unemployed, Disabled, Unskilled and Semi-Skilled 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 decomposition brings about 60% of low-skilled, 30% of high-skilled and 10% of farmers in 1881 (78% of low-skilled, 12% of high-skilled and 10% of farmers in 1817) in the 70 metropolitan areas. For 1971–2011, occupations are already classified into 1-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 between 1881 and 1971–2011. Clerks and Manual Skilled workers are thus classified as low-skilled, which brings about 62% of low-skilled and 38% of high-skilled in 2011.

1.3 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 towards 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 towards eastern neighborhoods, albeit with slightly greater noise.

[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).

[19] The left panel of Figure 6 is constructed as follows. First, we define a grid of equally-spaced points—every 100 meters—within 1.5 kilometers 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. The right panel is constructed in a same fashion with the share of low-skilled workers in 1881. We provide a similar illustration centered on town halls in Appendix Figure A4.

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 $\times$ chimney pairs where a block is a census cluster of households with the same geolocation in 1881 which is located within 2 kilometers of the chimney (there are 100,000 such blocks). The left panel of Figure 7 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 meters 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 the right panel of Figure 7, there is a 1.5 percentage point excess share of low-skilled workers for blocks situated north-east of the chimney. This gradient in the direction of prevailing winds is what will be captured in the reduced-form analysis of the next section.

Table 1 provides summary statistics at the level of our baseline units of observation. Within a buffer of 20 kilometers around the centroids of our 70 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 which 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 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 the last three columns 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.

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

2 Reduced-form evidence

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

2.1 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 influenced the location of pollution sources.

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

$$Y_{it} = \alpha + \beta P_i + \gamma \mathbf{X}_i + \nu \mathbf{Y}_p + \delta_c + \varepsilon_{ict} \tag{S1}$$

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 like 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 $\mathbf{X}_i$. $\mathbf{Y}_p$ is a set of measures of occupational structure in 1817 at the parish level (shares of low-skills, high-skills and farmers) and the logarithm of the property tax in 1815 at the parish level, which capture possible fixed neighborhood amenities. δ_c are city fixed effects 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

[21] Our main independent variable, i.e., historical pollution, is a pre-estimated regressor, which has possibly a sampling variance of its own, due to variation in the 10-year average weather conditions for instance. Standard errors need to be adjusted in such cases (Murphy and Topel, 2002). One standard procedure is to bootstrap the procedure defined by (1) the estimation of historical pollution and (2) our reduced-form regression. However, the sampling variance generated by step (1) occurs within the atmospheric dispersion model that we outsource to ADMS5. We thus cannot correct our standard errors for this issue, and they may be slightly under-estimated. 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.

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 Figure 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 wards.[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.

IV specification To account for the bias arising from potentially non-random 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 et al., 2012). We locate hypothetical chimneys in intervals of 150 meters along waterways in 1827, before the rise of coal as the 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 which 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

[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 which accounts for equilibrium adjustments at the city-level (see Section 3).

[23] Appendix 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) the resulting spatial distribution of air pollutants. Finally, panel (d) shows the distribution of air pollutants using actual pollution sources.

pollution, P_i, in Equation (S1):

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

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.

2.2 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 the first column, air pollution and the share of low-skilled workers in 1881 are positively correlated. Controlling for a large set of covariates does not affect the estimates. In the second column, we add city fixed-effects to control for variation in atmospheric pollution and neighborhood composition between cities (Hanlon, 2019). In the third column, 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. From the fourth to the last column, we add separate elements entering in the pollution dispersion process. In the fourth column, we condition on topography (elevation and distance to waterways in 1827). In the fifth column, we control for distance to pollution sources (heavy- and light-industry), distance to the city hall, distance to parks, area and the share of the LSOA within the 1880 city borders.[24] In the sixth column, we add eastings 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 (column 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

[24] As stated in Section 1, the metropolitan areas (and thus the sample of LSOAs) are defined by a buffer of 20 kilometers 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).

[25] The first column of Appendix Table A2 reports the coefficients on all covariates.

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 (respectively in 1817 and in 1881) 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, i.e., 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 symmetric in all directions. We then generate a set of counterfactual pollution exposures using wind profiles *rotated* in steps of 30 degrees 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 (Panel 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 (Panel b), after pollution became a meaningful disamenity, we see a pronounced, bell-shaped pattern with a peak in correlation observed around wind profiles ro-

[26] We consider other outcomes in Appendix Table A3, with the share of all low-skilled workers including females and the share of migrants, distinguishing between migrants from England and Wales and from 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 (thus unrelated to the Irish Potato Famine).

[27] The exercise underlying Figure 9 is similar in nature to the exercise underlying Figure 7. In both cases, the objective is to exploit the direction of prevailing winds and the location of neighborhoods relative to chimneys, controlling for their proximity to such chimneys. Relative to Figure 7, Figure 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 column 6 of Table 2).

tated by 0 and 30 degrees.[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 Panel (c) and for residential pollution in Panel (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 degrees.

Finally, we present in Table 3 the results of the IV strategy (S2), that uses 1827 waterways as a source of exogenous variation for chimney locations. We report four sets of estimates, excluding neighborhoods within 250m or 500m of waterways, and with or without extended controls in addition to city fixed-effects. The first stage is strong in all cases, and the (stable) 2SLS 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 standard deviations above the mean (see Figure 8); the average pollution for central neighborhoods or for neighborhoods in proximity to waterways is around 0.10-0.30. Moreover, we find larger average treatment effects when restricting the sample to central LSOAs (see Appendix E).

2.3 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, as estimated by Equation (S1). One standard deviation in historical air pollution increases the prevalence of low-skilled workers by 2.5 to 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

[28] The fact that the peak in correlation is between 0 and 30 degrees may be due to measurement error. First, wind patterns may have changed in one century, in particular the frequency of cyclonic or anti-cyclonic 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.

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 in England and Wales as recorded by: (i) the Land Registry between 2000 and 2011 (columns 1 and 2); and, (ii) Nationwide Building Society between 2009 and 2013 (columns 3 and 4). Hedonic regressions with and without controlling for average house characteristics find that one additional standard deviation in past pollution is associated with a price drop of about 10% to 11%. Controlling for property characteristics reduces these estimates to 6-8%, showing that properties in formerly polluted neighborhoods are smaller and more likely to be "non-detached".[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 non-linearities 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 one standard deviation 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 non-linearities, we define ten pollution categories corresponding to the ten deciles in intra-city pollution, i.e., 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 the left panel of Figure 11, while the estimates for pollution-decile trends are shown in the right panel. 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. The right

[29] Appendix 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

panel of Figure 11 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 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.

2.4 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 pre-existing 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 only vary the chimney height as the exit velocity and the exit temperature would affect the same crucial model input, i.e., the height of the smoke column in the atmosphere. The atmospheric pollution based on chimneys assumed to be shorter (15m) or taller (40m) than our baseline (25m) 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 (i) the North-East, (ii) the North-West, (iii) the South-East, or (iv) the South-West quadrants—within a given distance from each grid point. Lastly, we collapse the measure for all four quadrants into 2001 LSOAs. Only chimneys located North-West and South-West 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 South-West quadrants are relatively more predictive of deprivation in Northern

$\theta < 1$ the AR(1) parameter, we should observe that:

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

cities, reflecting the more southerly wind direction in North England (see Figure 4).

Third, we test the sensitivity of our findings to the addition of the following pollution imprints as controls: a 'symmetric' placebo pollution measure which assumes that pollution spreads evenly in all directions and captures the proximity to industrial centers; a placebo pollution pattern which 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 between 3.6 and 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. 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 North-West; the North-East. 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.

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

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 which would affect disproportionately downwind neighborhoods. To predict early industrial districts, we locate 543 early steam engines installed between 1700–1800, using data from Kanefsky and Robey (1980) and Nuvolari et al. (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 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.

3 A dynamic model of residential sorting

In order to quantify the non-linear 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 4.

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 with $a_{j,t}$. 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 as follows:

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

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 non-movers 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:

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

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 $\left\{a_{j,t}\right\}_{j,t}$ as given, and maximizes:

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

We assume, as common in models of residential sorting (Bayer et al., 2016), that idiosyncratic preferences, $\varepsilon_{i,j,t}$, are distributed across households along a Type 1 Extreme-Value distribution such that the fraction of households opting for neigh-

[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.

borhood 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 only need to consider the states of nature along which no further re-optimization 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.

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

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

[33] We allow for this share of workers of different skill types to fluctuate over time, following migration choices, in Appendix 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 Appendix 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.

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]

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

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 which 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 that 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

[34] The reader interested in the detailed derivation of the demand schedule can refer to Appendix 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.

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.

4 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.

Identification We estimate Equation (2) using the observed neighborhood composition in 142 closed cities, indexed by c, over five waves (1971, 1981, 1991, 2001, 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 only differ along their fixed income y. Specifically, we let the utility of a type-o household living in neighborhood j of city c be,

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

We further assume that the neighborhood amenity, $a_{j,c,t} = h\left(a_{j,c}, \chi_{j,c,t}, s_{j,c,t}\right)$, is a function of (i) a constant, city-wide exogenous amenity, $a_{j,c}$, (ii) a neighborhood amenity shock, $\chi_{j,c,t}$, satisfying $E_{t-1}\left[\chi_{j,c,t}\right] = 0$ 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 Bayer et al., 2016, for instance), even at the aggregate level. Therefore, we can only identify a function h which is separable in its components. We show in Appendix B.2 that demand for neighborhood j can then be written as,

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

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

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, $F\left(\frac{s_{j,c,t+1}-\theta s_{j,c,t}}{1-\theta}\right) - E_t F\left(\frac{s_{j,c,t+1}-\theta s_{j,c,t}}{1-\theta}\right)$, thus satisfying $E_{t-1}\left[\zeta_{j,c,t}\right] = 0$.

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 Equation (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 crucial in separating the specific role of relocation rigidities. The following thought experiment helps understand this last argument. Consider one neighborhood which 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 they 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 Equation (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 allow 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 Figures 10 and 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

[37] We study the dynamic properties of the system of Equations (S3) in Appendix B.4.

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 non-linear dynamics linked to the presence of amenities in a neighborhood by interacting a measure of predicted amenities (using all controls of Table 2, column 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 only affects 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.

Structural estimates The estimates for Equation (S3) are reported in Table 7. The first column reports the baseline specification where neighborhood preferences are linear in neighborhood composition. The share of low-skilled workers 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 Appendix 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 ten 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 neighbourhoods), 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

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

exploit the non-linearities in neighborhood dynamics documented in Figure 11 (see column 2, Table 7). 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 (column 4). Finally, we estimate a non-linear 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.

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 we adjust the neighborhood fixed amenity to keep the right-hand side of Equation (S3) unchanged in 1971. In panel (a) of Appendix Figure A9, 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 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 panel (b) of Appendix Figure A9. 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.

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 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 (columns 1 and 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 Figure 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 blue dots represent cities in the top quartile of pollution; light blue dots represent cities in the lower quartiles of pollution. In 1971, just after the Clean Air Act of 1968, the *Dissimilarity* measure is about .07 (about one standard deviation) 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.

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,

[39] The *Dissimilarity* measure is defined as $\frac{1}{2J_c}\sum_{j=1}^{J_c}\left|\frac{s_{j,c}}{s_c} - \frac{1-s_{j,c}}{1-s_c}\right|$ 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.

and we report the correlation between past pollution and the simulated deprivation in panel (a) of Figure 14. 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 panel b of Figure 14).

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 one standard deviation lower with locational rigidities for those cities with an above-median stock of social housing in 1971. The reason for this counter-intuitive finding is that social housing is negatively correlated with past pollution in 1971 and only aligns with deprivation from 1991 onward (see Appendix G). A policy which would confine social housing to its pre-determined location would thus limit the extent to which well-connected, attractive locations would further gentrify (as documented, for instance, in Guerrieri et al., 2013; Baum-Snow and Hartley, 2020; Couture and Handbury, 2019).

5 Conclusion

This paper presents a plausible explanation for what was, until this paper, 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 the industrialization, pollution became a major environmental disamenity in cities. An unequal distribution of pollution exposure induced a sorting process which 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 non-linear transitional dynamics which relate to the literature on tipping dynamics (Card et al., 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 like 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 an uneven pollution exposure across space: pollution induces spatial inequalities that far outlive de-industrialization.

References

Anderson, Michael L., "As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality," *Journal of the European Economic Association*, 2019, *forthcoming.*

Banzhaf, H. Spencer and Randall P. Walsh, "Do People Vote with Their Feet? An Empirical Test of Tiebout," *American Economic Review*, June 2008, *98* (3), 843–63.

Barreca, Alan, Karen Clay, and Joel Tarr, "Coal, Smoke, and Death: Bituminous Coal and American Home Heating," Working Paper 20648, National Bureau of Economic Research February 2014.

Baum-Snow, Nathaniel and Daniel A. Hartley, "Accounting for central neighborhood change, 1980–2010," *Journal of Urban Economics*, 2020, *117.*

Bayer, Patrick, Robert McMillan, Alvin Murphy, and Christopher Timmins, "A dynamic model of demand for houses and neighborhoods," *Econometrica*, 2016, *84* (3), 893–942.

Beach, Brian and W. Walker Hanlon, "Coal Smoke and Mortality in an Early Industrial Economy," *Economic Journal*, 2018, *128*, 2652–2675.

Bleakley, Hoyt and Jeffrey Lin, "Portage and Path Dependence," *The Quarterly Journal of Economics*, 2012, *127* (2), 587–644.

Brimblecombe, Peter, "The Clean Air Act after 50 years," *Weather*, 2006, *61* (11), 311–314.

Brooks, Leah and Byron Lutz, "Vestiges of Transit: Urban Persistence at a Microscale," *Review of Economics and Statistics*, 2019, *101* (3), 385–399.

Brueckner, Jan K and Stuart S Rosenthal, "Gentrification and neighborhood housing cycles: will America's future downtowns be rich?," *The Review of Economics and Statistics*, 2009, *91* (4), 725–743.

Card, David, Alexandre Mas, and Jesse Rothstein, "Tipping and the Dynamics of Segregation," *The Quarterly Journal of Economics*, 2008, *123* (1), 177–218.

Chay, Kenneth Y. and Michael Greenstone, "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy*, 2005, *113* (2), 376–424.

Chen, Shuai, Paulina Oliva, and Peng Zhang, "The Effect of Air Pollution on Migration: Evidence from China," Working Paper 24036, National Bureau of Economic Research November 2017.

Clark, G. Kitson, *The Making of Victorian England*, Harvard University Press, 1962.

Clay, Karen, Joshua Lewis, and Edson Severnini, "Canary in a Coal Mine: Infant Mortality, Property Values, and Tradeoffs Associated with Mid-20th Century Air Pollution," Technical Report 22155 April 2016.

Conley, Timothy G, "GMM estimation with cross sectional dependence," *Journal of Econometrics*, 1999, *92* (1), 1–45.

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

_ , "Records of industrial melanism in British moths," *Biological Journal of the Linnean Society*, 2018, *125*, 862–866.

Couture, Victor and Jessie Handbury, "Urban Revival in America, 2000 to 2010," Working Paper, National Bureau of Economic Research 2019.

Crafts, N. F. R. and C. K. Harley, "Output Growth and the British Industrial Revolution: A Restatement of the Crafts-Harley View," *The Economic History Review*, 1992, *45* (4), 703–30.

Crafts, Nicholas, "Economic growth during the long twentieth century. Chapter 2 in Floud, R., Humphries, J. and Johnson, P.," 2014. Cambridge University Press.

Davis, Donald R and David E Weinstein, "Bones, bombs, and break points: the geography of economic activity," *American Economic Review*, 2002, *92* (5), 1269–1289.

Depro, Brooks, Christopher Timmins, and Maggie O'Neil, "White Flight and Coming to the Nuisance: Can Residential Mobility Explain Environmental Injustice?," *Journal of the Association of Environmental and Resource Economists*, 2015, *2* (3), 439–468.

Deryugina, Tatyana, Garth Heutel, Nolan H. Miller, David Molitor, and Julian Reif, "The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction," *American Economic Review*, 2019, *109* (12), 4178–4219.

Durlauf, Steven N., "A Theory of Persistent Income Inequality," *Journal of Economic Growth*, 1996, *1*, 75–93.

_ , "Neighborhood Effects," in J. Vernon Henderson and J.F. Thisse, eds., *Handbook of Regional and Urban Economics*, Vol. 4, Elsevier, 2004, pp. 2173–2242.

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

Graff Zivin, Joshua and Matthew Neidell, "Environment, Health, and Human Capital," *Journal of Economic Literature*, 2013, *51* (3), 689–730.

Guerrieri, Veronica, Daniel Hartley, and Erik Hurst, "Endogenous Gentrification and Housing Price Dynamics," *Journal of Public Economics*, 2013, *100*, 45–60.

Hanlon, W. Walker, "Coal Smoke and the Costs of the Industrial Revolution," *Economic Journal*, February 2019, *130*, 462–488.

Heblich, Stephan, Stephen J. Redding, and Daniel M. Sturm, "The Making of the Modern Metropolis: Evidence from London," *Quarterly Journal of Economics*, 2020, (forthcoming).

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

Hornbeck, Richard and Daniel Keniston, "Creative Destruction: Barriers to Urban Growth and the Great Boston Fire of 1872," *American Economic Review*, 2017, *107*, 1365–1398.

Kanefsky, John and John Robey, "Steam Engines in 18th-Century Britain: A Quantitative Assessment," *Technology and Culture*, 1980, *21* (2), 161–186.

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

Kuminoff, Nicolai V., V. Kerry Smith, and Christopher Timmins, "The new economics of equilibrium sorting and policy evaluation using housing markets," *Journal of Economic Literature*, 2013, *51* (4), 1007–1062.

Lamb, HH, "British Isles Weather types and a register of daily sequence of circulation patterns. Geophysical Memoir 116," 1972.

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

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

Lee, Sanghoon and Jeffrey Lin, "Natural Amenities, Neighborhood Dynamics, and Persistence in the Spatial Distribution of Income," *The Review of Economic Studies*, 2018, *85*, 663–694.

Logan, Trevon and John M. Parman, "The National Rise in Residential Segregation," *Journal of Economic History*, 2017, *77* (1), 127–170.

Maw, Peter, Terry Wyke, and Alan Kidd, "Canals, rivers, and the industrial city: Manchester's industrial waterfront, 1790–1850," *The Economic History Review*, 2012, *65* (4), 1495–1523.

Monras, Joan, "Immigration and wage dynamics: Evidence from the Mexican Peso Crisis," *Journal of Political Economy*, 2020, (forthcoming).

Mosley, Stephen, *The chimney of the world: a history of smoke pollution in Victorian and Edwardian Manchester*, Routledge, 2013.

Murphy, Kevin M and Robert H Topel, "Estimation and inference in two-step econometric models," *Journal of Business & Economic Statistics*, 2002, *20* (1), 88–97.

Musson, A. E., "Industrial Motive Power in the United Kingdom, 1800-70," *The Economic History Review*, 1976, *29* (3), 415–39.

Negro, Marco Del, Domenico Giannone, Marc P Giannoni, and Andrea Tambalotti, "Global trends in interest rates," *Journal of International Economics*, 2019, *118*, 248–262.

Nuvolari, Alessandro, Bart Verspagen, and Nick von Tunzelmann, "The early diffusion of the steam engine in Britain, 1700–1800: a reappraisal," *Cliometrica*, 2011, *5* (3), 291–321.

Ravenstein, E. G., "The laws of migration," *Journal of the Statistical Society of London*, 1885, *48*, 167–235.

Redding, Stephen J. and Daniel M. Sturm, "Estimating Neighborhood Effects: Evidence from War-time Destruction in London," Technical Report 2016.

Rosenthal, Stuart S., "Old homes, externalities, and poor neighborhoods. A model of urban decline and renewal," *Journal of Urban Economics*, 2008, *63*, 816–840.

_ **and Stephen Ross**, "Change and Persistence in the Economic Status of Neighborhoods and Cities," in J. Vernon Henderson Gilles Duranton and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5 of *Handbook of Regional and Urban Economics*, Elsevier, 2015, pp. 1048–1120.

Schelling, Thomas C, "Dynamic models of segregation," *Journal of Mathematical sociology*, 1971, *1* (2), 143–186.

Shaw-Taylor, L. and E. A. Wrigley, "Occupational structural and population change. Chapter 2 in Floud, R., Humphries, J. and Johnson P.," *The Cambridge Economic History of Modern Britain: Volume 1, Industrialisation 1700-1860*, 2014. Cambridge University Press.

Siodla, James, "Razing San Francisco: The 1906 disaster as a natural experiment in urban redevelopment," *Journal of Urban Economics*, 2015, *89*, 48–61.

Villarreal, Carlos, "Where the Other Half Lives: Evidence on the Origin and Persistence of Poor Neighborhoods from New York City 1830-2012," Technical Report January 2014.

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

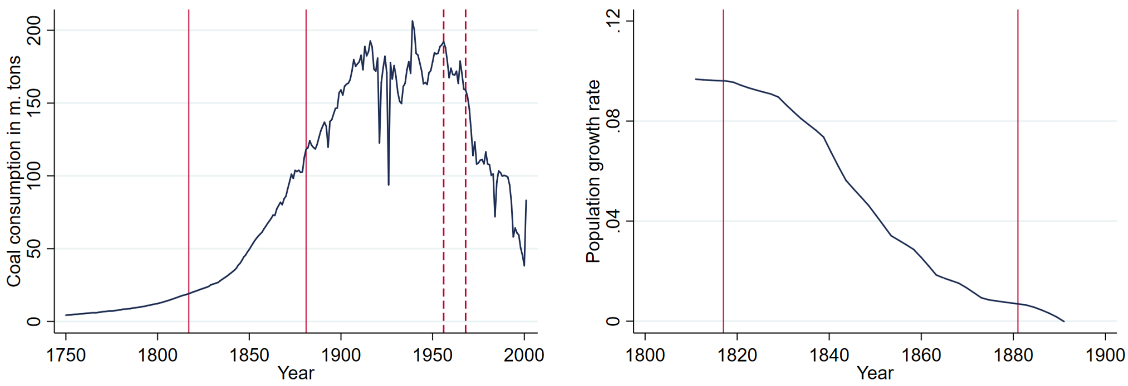
Williamson, J. G., *Coping with City Growth during the British Industrial Revolution*, Cambridge University Press, 1990.

Wrigley, E. A., "The PST System of Classifying Occupations," 2010.

Zivin, Joshua Graff and Matthew Neidell, "The Impact of Pollution on Worker Productivity," *American Economic Review*, December 2012, *102* (7), 3652–73.

Figures and tables

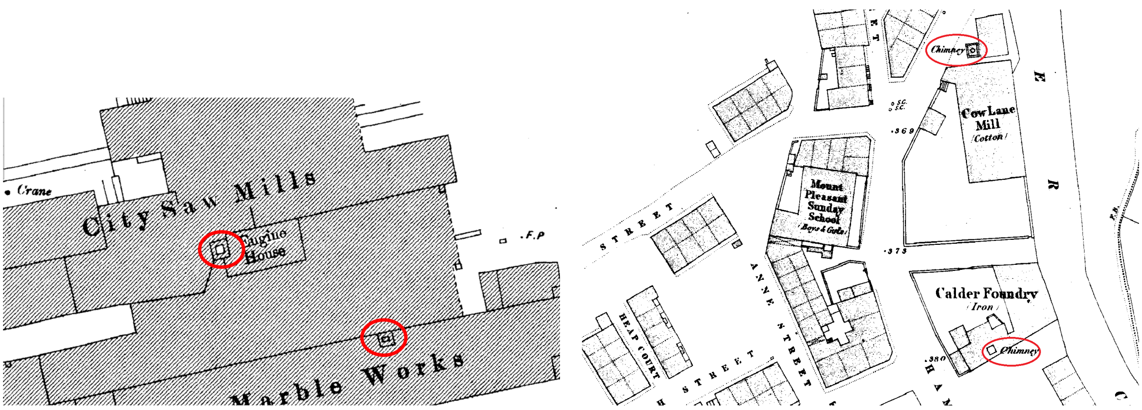
Figure 1. Coal consumption and migration during the Industrial Revolution.



(a) Coal consumption (1750–2001) (b) Urban growth rates (1801–1891)

Notes: The left panel illustrates the increase and decrease in coal consumption over the period 1750–2000. The figure is based on 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 red lines indicate the years 1817 and 1881, while the dashed red lines mark the introduction of the 1956 and 1968 Clean Air Acts. The right panel plots the average decadal population growth rate for the period 1801–1891 in cities of our sample.

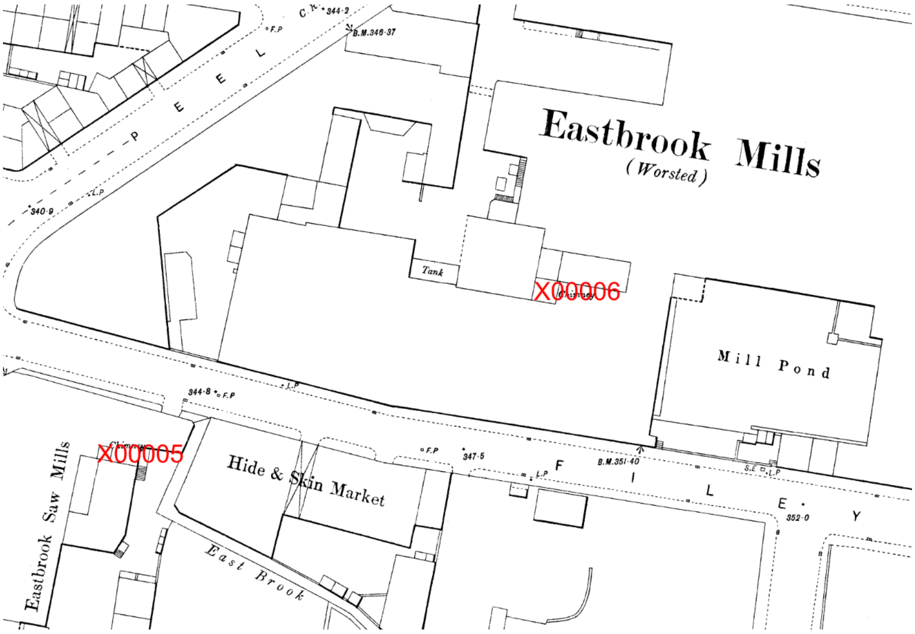
Figure 2. Ordnance Survey maps—chimney symbols.



(a) Example 1. (b) Example 2.

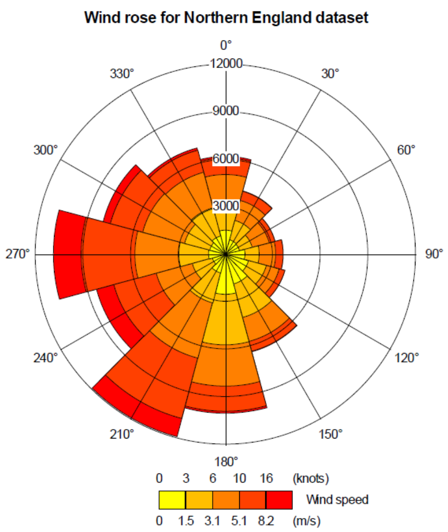
Sources: Ordnance Survey Maps—25 inch to the mile, 1842–1952. Four different symbols for chimneys are circled. The variations in symbols prevent us from directly using a recognition algorithm. Instead, we go through all maps and mark chimneys with a red symbol X and a unique numeric identifier (see Figure 3).

Figure 3. Town maps—marking and identifying chimneys.

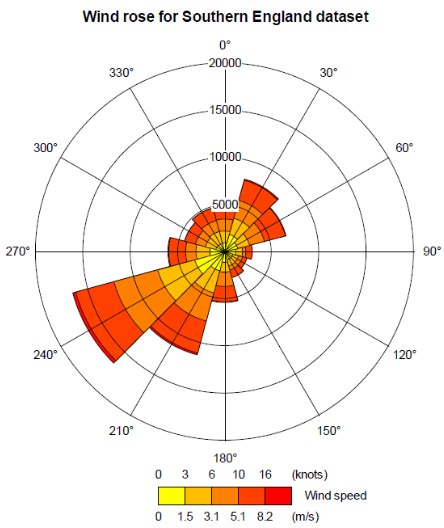


Sources: Ordnance Survey Maps—25 inch to the mile, 1842–1952. Marks X and the identifiers, e.g., *00006*, are used by a recognition algorithm to locate chimneys and associate a factory. Together with the projection provided by the Ordnance Survey, allows us to geolocate each chimney. Factory-specific information can be retrieved after the recognition algorithm has (i) located a chimney and (ii) stored the associated identifier.

Figure 4. Wind roses differences across two sets of meteorological conditions.



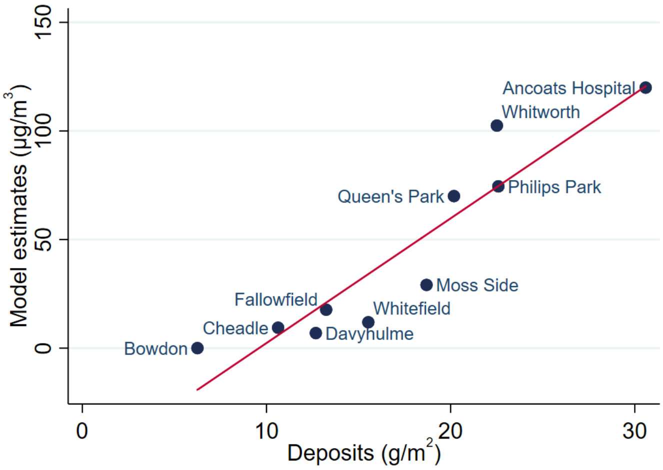
(a) North England.



(b) South England.

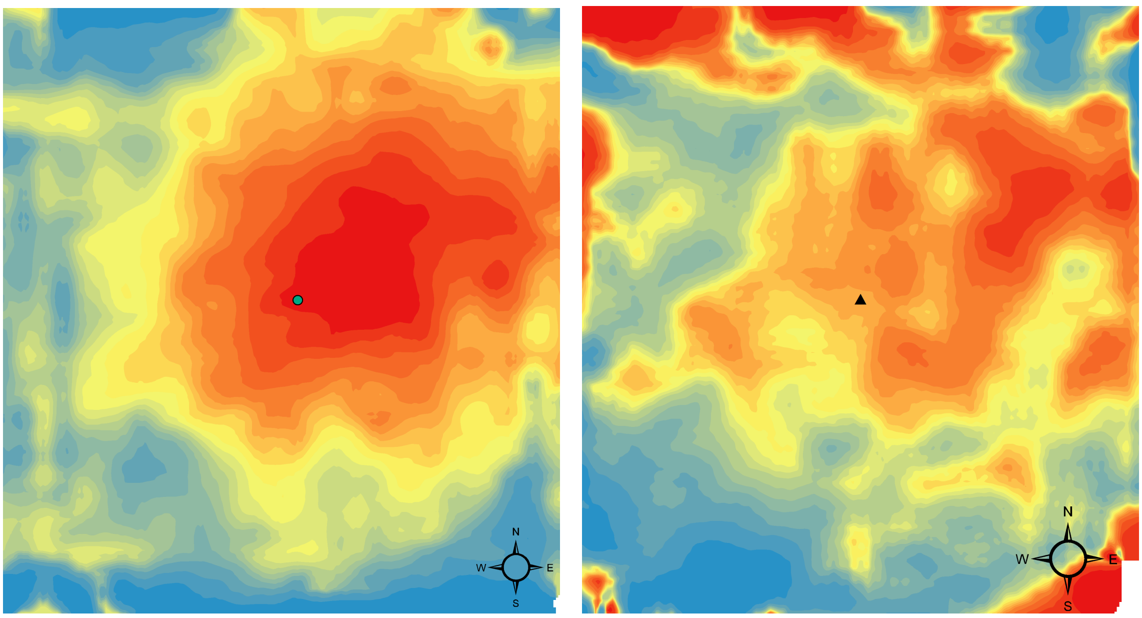
Sources: Met Office—10-year statistical meteorological data. We use 4 different sets of meteorological conditions across England and Wales: Southern England, Central England, Northern England and East Anglia.

Figure 5. Air pollution measures across neighborhoods of Manchester—external validity.



Source: First Annual Report of the Sanitary Committee on the Work of the Air Pollution Advisory Board, 1915. This Figure reports the relationship between deposits, as collected by the Air Pollution Advisory Board (1915), and our measure of $SO2$ concentration in $\mu g/m^3$. 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), Bowdon (6.25).

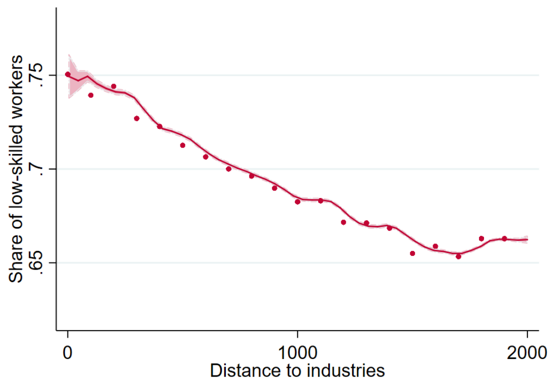
Figure 6. Pollution and shares of low-skilled workers in the average city.



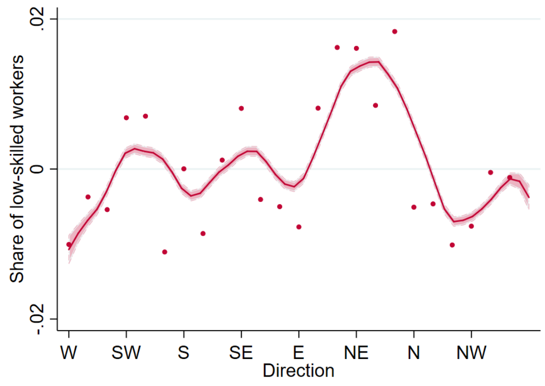
(a) Pollution (1880–1900). (b) Share of low-skilled workers (1881).

Notes: The left panel of this Figure displays the average gradient of pollution in 1880–1900 across cities. The right panel displays the 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 kilometers 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 colors/level lines are quantiles of pollution (left panel) and share of low-skilled workers (right panel); red indicates the highest quantiles.

Figure 7. Shares of low-skilled workers and position relative to pollution sources.



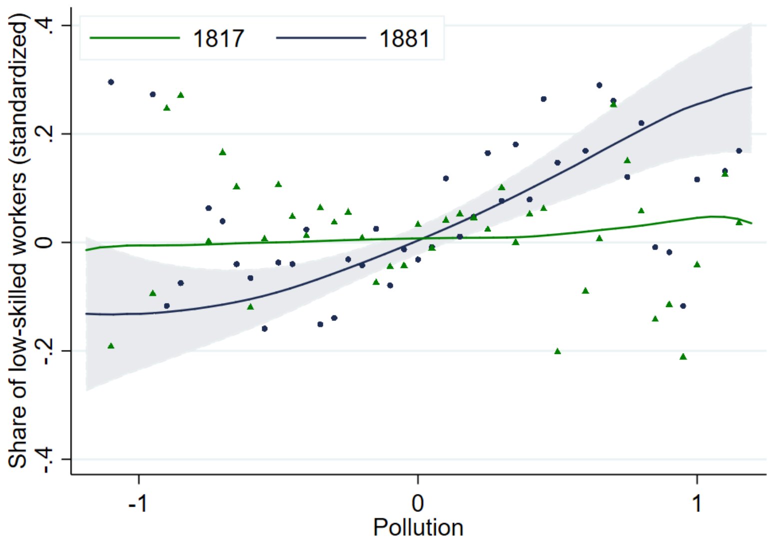
(a) Distance.



(b) Direction.

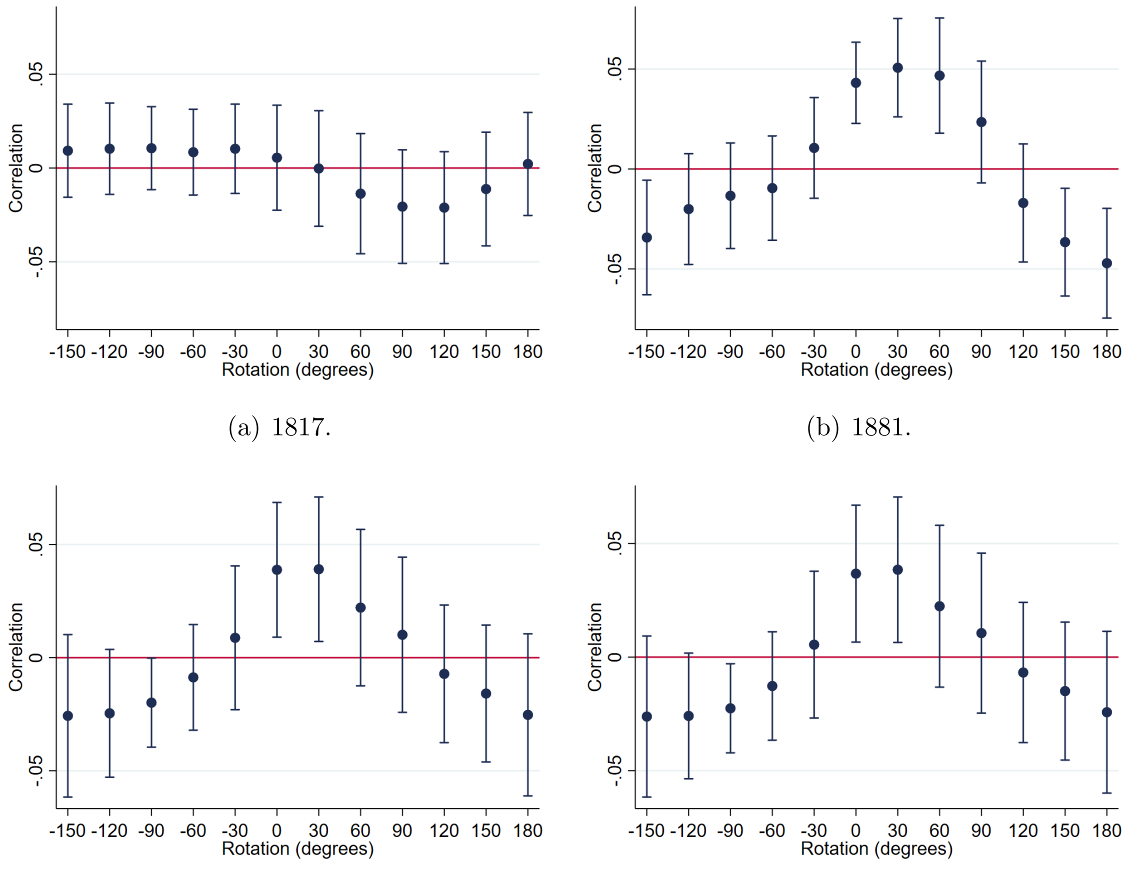
Notes: This Figure represents the relationship between the share of low-skilled workers in 1881 and the position relative to a pollution source. In this exercise, the unit of observation is a neighborhood × chimney pair where a neighborhood is a Census cluster of households with the same geolocation in 1881 (about 100 households). The left panel displays the average share of low-skilled workers in 1881 across observed units (weighted such that all households are given the same weight). The right panel displays the residual of the share of low-skilled workers in 1881 cleaned for distance to the pollution source and distance to amenities in 1881 (distance to 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 meters (panel a) and 20 degrees (panel b). *NE* stands for north-east, indicating that the household is located toward the north-east direction, from the standpoint of the pollution source.

Figure 8. Pollution (x-axis) across neighborhoods and shares of low-skilled workers (y-axis) in 1817 and 1881.



Notes: This Figure represents the relationship between the (standardized) shares of low-skilled workers in 1817 (teal, triangles) and 1881 (blue, circles) 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 *lat./lon.* controls (see Table 1). We create 40 bins of neighborhoods along past pollution and the dots 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(s).

Figure 9. Rotating wind patterns in 1817 and 1881.



(a) 1817.

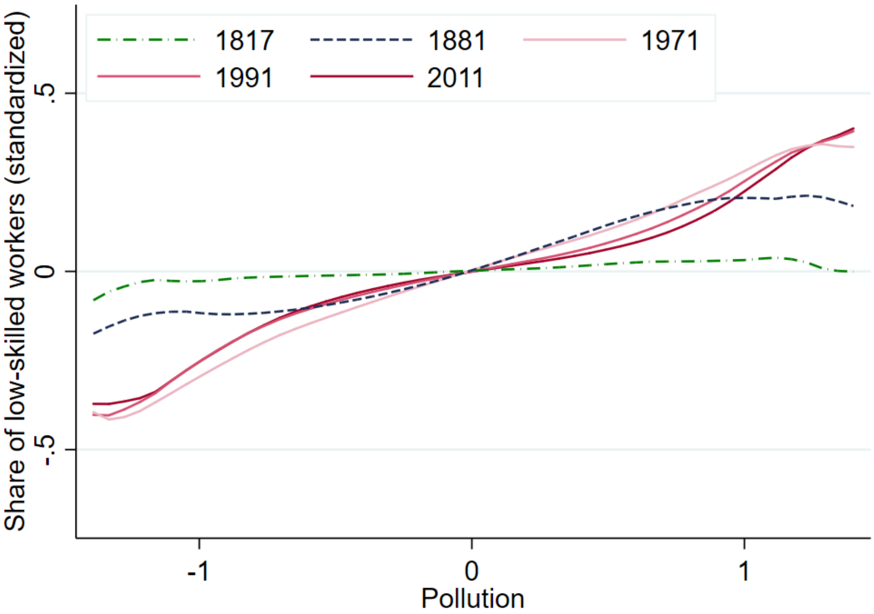
(b) 1881.

(c) 1881 (parish FEs).

(d) 1881 (parish FEs and residential pollution).

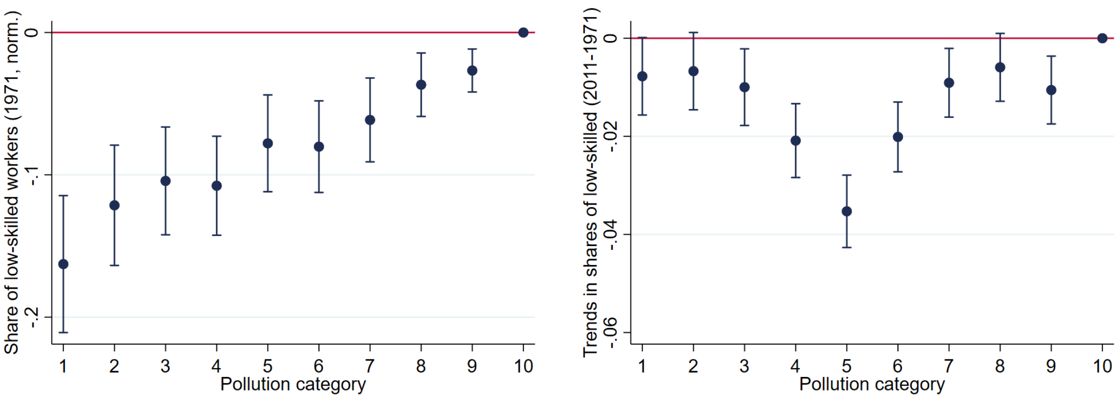
Notes: These Figures represent the conditional correlations between the shares of low-skilled workers (in 1817 and in 1881) and counterfactual measures of past pollution rotated in steps of 30 degrees. Each dot represents the estimate in a specification including the controls reported in Table 1, Column 6, and the measure of *symmetric pollution* capturing proximity to the pollution source. In Panel (c), we control for parish fixed effects. In Panel (d), we control for parish fixed effects and residential pollution. Standard errors are clustered at the parish-level, and the lines represent 5% confidence intervals.

Figure 10. Pollution (x-axis) across neighborhoods and shares of low-skilled workers (y-axis) in 1817, 1881, 1971, 1991 and 2011.



Notes: 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, and *topography* and *population* controls.

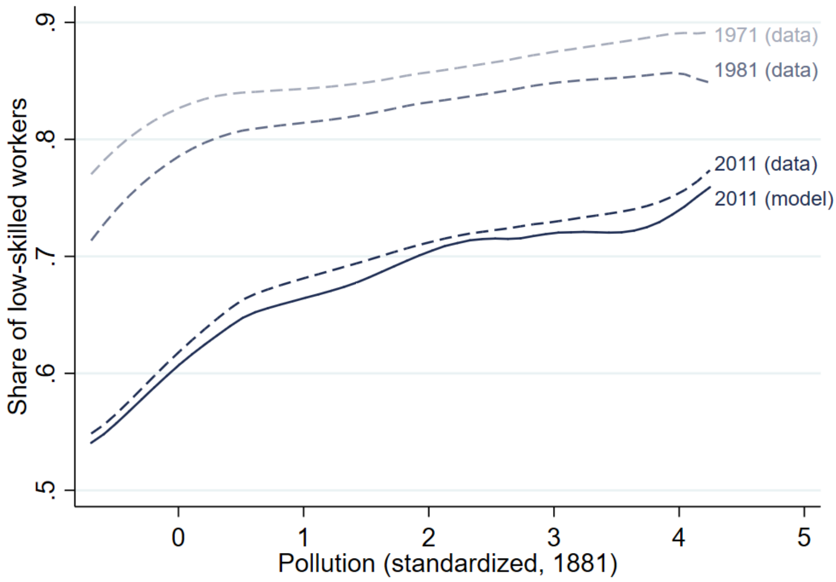
Figure 11. Mean reversion between 1971 and 2011—share of low-skilled workers by pollution decile.



(a) 1971 (b) Decadal change between 1971 and 2011

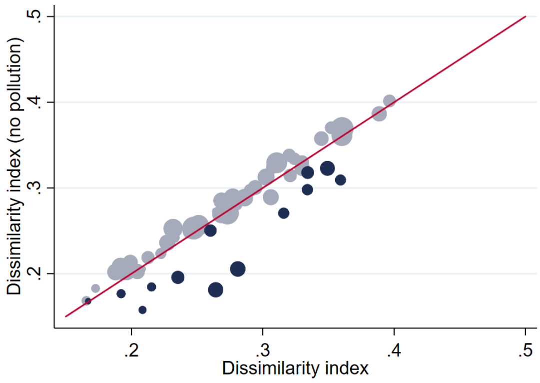
Notes: The left panel represents the effect of pollution on the share of low-skilled workers in 1971 in a specification with pollution category dummies defined by pollution decile (1: lowest, 10: highest pollution exposure). The effect of the last decile is normalized to 0. The right panel represents 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, 2011 with LSOA fixed effects. The reported coefficients are extracted from the interaction of pollution category dummies and the year, and adjusted such as to represent decadal changes.

Figure 12. Simulated dynamics (1971–2011) using the structural estimates.

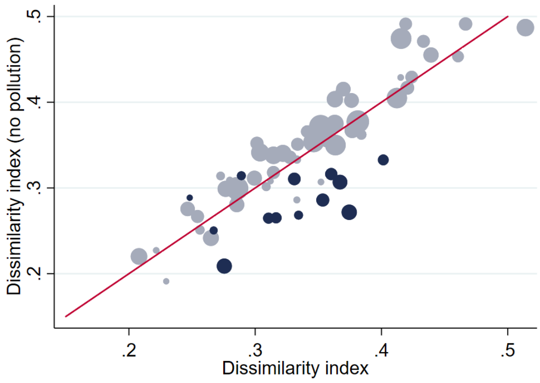


Notes: This Figure shows the average shares of low-skilled workers as a function of past pollution, using a local polynomial smoothing. We report neighborhood dynamics, as estimated using the structural estimates reported in column 1 of Table 6 and the observed shares of low-skilled workers in 1971 and 1981. Dashed lines (1971, 1981, 2011 from lighter to darker blue) represent observed shares of low-skilled workers; plain lines represent simulated shares in 2011.

Figure 13. Neighborhood segregation across cities in the absence of pollution.



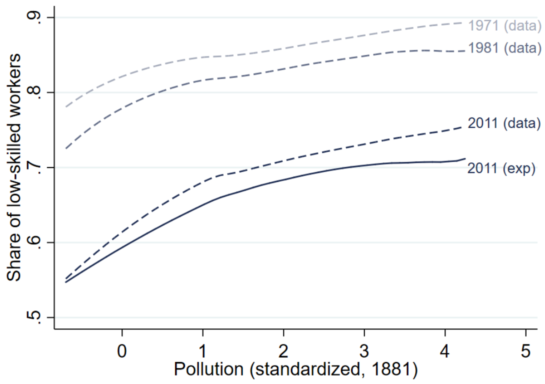
(a) No pollution (1971)



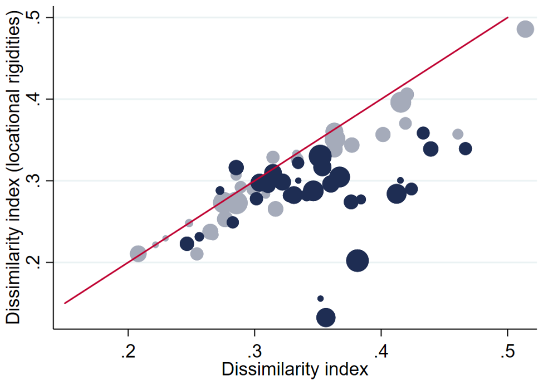
(b) No pollution (2011)

Notes: This Figure compares measures of segregation across cities in two experiments: (i) the simulated dynamics based on actual neighborhood compositions in 1971 and 1981 (x-axis) and the 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 (columns 1 and 2). Each city is represented by a dot whose size is proportional to its number of LSOAs within the sample and whose color shows the importance of pollution at the city level. Dark blue dots represent cities in the top quartile of pollution; light blue dots represent cities in the lower quartiles of pollution. The measure of segregation used is a measure of evenness, i.e., dissimilarity: $\frac{1}{2J_c}\sum_{j=1}^{J_c}\left|\frac{s_{j,c}}{s_c} - \frac{1-s_{j,c}}{1-s_c}\right|$ 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.

Figure 14. Neighborhood dynamics and segregation with fixed social housing.



(a) Simulated dynamics



(b) Neighborhood segregation

Notes: In panel (a), we report 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 LSOA level (corresponding to the minimum between this share and fractions of low-skilled workers in 1971 and 1981). Panel (b) compares measures of segregation across cities in two experiments: (i) the unrestricted simulated dynamics and (ii) the simulated dynamics with locational rigidities (y-axis). Each city is represented by a dot 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 blue dots represent cities with above-median shares of social housing in 1971; light blue dots represent cities with below-median shares of social housing in 1971. The measure of segregation used is a measure of evenness, i.e., dissimilarity: $\frac{1}{2J_c}\sum_{j=1}^{J_c}\left|\frac{s_{j,c}}{s_c} - \frac{1-s_{j,c}}{1-s_c}\right|$ 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.

Table 1. Descriptive statistics and variance decomposition.

		Pollution		Standard deviation		
VARIABLES	Mean	high	low	total	between	within
		Air pollution				
Normalized pollution	-.034	.234	-.298	.928	.596	.560
		Population measures				
1817†						
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
		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
		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 (square km)	.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

Notes: All statistics are computed using the baseline sample of 5,538 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. † Shares in 1817 are computed at the parish-level, which explains the lower variance.

Table 2. Pollution and shares of low-skilled workers in 1881.

Share of low-skilled	(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
Controls (topography)	No	No	No	Yes	Yes	Yes
Controls (amenities)	No	No	No	No	Yes	Yes
Controls (lat./lon.)	No	No	No	No	No	Yes

Standard errors are reported between parentheses and are clustered at the parish-level. Standardized effects are reported between square brackets. Each cell is the result of a separate regression. The unit of observation is a Lower Super Output Area. 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, the (inverse) distance to parks, the share of LSOA within the city borders in 1880, the LSOA area, the (inverse) distance to the closest heavy industry and the (inverse) distance to the closest light industry. *Lat./lon.* are the latitude and longitude of the LSOA centroid.

Table 3. Pollution and shares of low-skilled workers in 1881—IV specification.

Panel A: First stage	Pollution			
	(1)	(2)	(3)	(4)
Pollution (waterways)	.2904	.2020	.3100	.1834
	(.0331)	(.0347)	(.0403)	(.0384)

Panel B: Second stage	Share of low-skilled workers			
	(1)	(2)	(3)	(4)
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>250m	Canal>250m	Canal>500m	Canal>500m
Fixed effects (city)	Yes	Yes	Yes	Yes
Extended controls	No	Yes	No	Yes

Standard errors are reported between parentheses and are clustered at the parish-level. Standardized effects are reported between square brackets. The top panel reports the first stage, and Kleibergen-Paap F-statistics are reported in the bottom panel. The unit of observation is a Lower Super Output Area. The set of extended controls include all controls of column 6 in Table 2. The variable *Pollution (waterways)* is the first predicted pollution instrument from a uniform allocation of pollution sources along waterways (as of 1827). In columns 1 and 2 (resp. 3 and 4), we exclude LSOAs within 250 meters (resp. 500 meters) of a waterway.

Table 4. Pollution and shares of low-skilled workers in 1971–2011.

Share of low-skilled workers	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

Standard errors are reported between parentheses and are clustered at the parish-level (as defined in 1881). Standardized effects are reported between square brackets. Each cell is the result of a separate regression. The unit of observation is a Lower Super Output Area. The set of extended controls include all controls of column 6 in Table 2.

Table 5. Pollution and shares of low-skilled workers in 1971–2011—IV specification.

Share of low-skilled workers	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

Standard errors are reported between parentheses and are clustered at the parish-level (as defined in 1881). Standardized effects are reported between square brackets. Each cell is the result of a separate regression. The unit of observation is a Lower Super Output Area. The set of extended controls include all controls of column 6 in Table 2. As in Table 3–columns 1 and 2, the instrument is the predicted pollution generated by a uniform allocation of pollution sources along waterways (as of 1827), and we exclude LSOAs within 250 meters of a waterway.

Table 6. Pollution, house prices and transactions (Nationwide, 2009–2013, and Land registry, 2000–2011).

	Nationwide		Land registry	
House prices	(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 ch.)	No	Yes	No	Yes

Standard errors are reported between parentheses and are clustered at the parish-level. Standardized effects are reported between square brackets. Each column is the result of a separate regression. The unit of observation is a Lower Super Output Area. The set of extended controls include all controls of column 6 in Table 2. The dependent variables are the (log) average house prices (from Nationwide in columns 1 and 2, and Land registry in columns 3 and 4). The unreported effect of pollution on the number of transactions is between -0.0473 (specification similar to column 3) and -0.0878 (specification similar to column 4). In column 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 the Nationwide transactions. In column 4, controls for house characteristics include the average shares of detached, semi-detached, terraced houses and new houses for all transactions.

Table 7. Relative demand for neighborhoods—structural estimation.

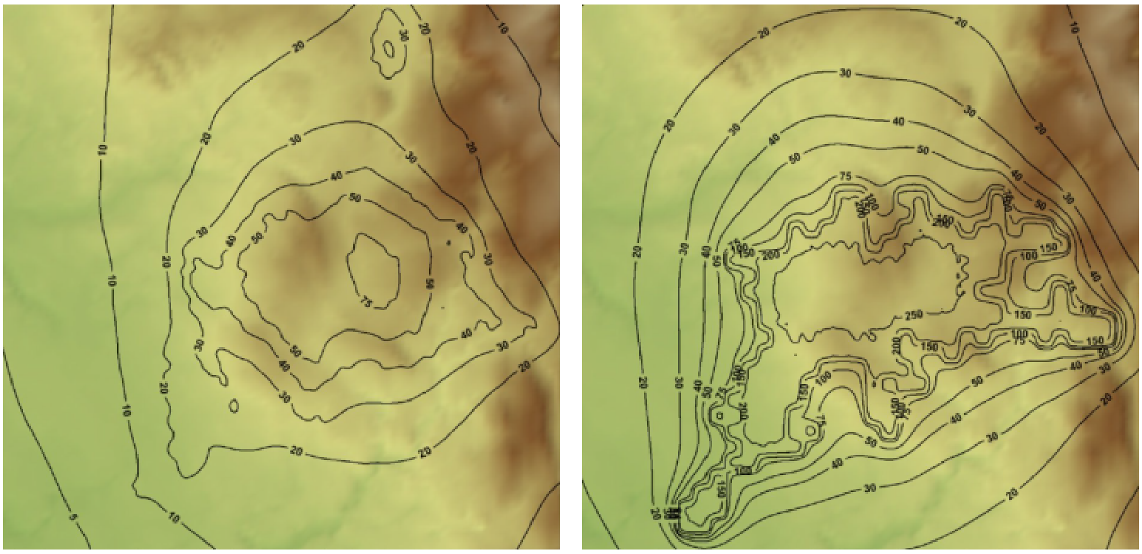
Specification (S3)	(1)	(2)	(3)	(4)	(5)
Relocation rigidity θ	.4782	.4777	.4706	.4730	.4897
	(.0229)	(.0221)	(.0314)	(.0209)	(.0225)
Discount factor β	.9303	.9000	.6312	.8171	1.365
	(.3783)	(.3152)	(.9056)		(.2660)
Preferences α_1	-2.869	-2.893	-3.652	-3.122	-5.285
	(1.051)	(.9147)	(2.062)	(.7225)	(1.590)
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)

Standard errors are reported between parentheses, and clustered at the parish $\times$ wave level. The unit of observation is a LSOA 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 which are replaced by measures of historical pollution interacted with 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, (iii) a measure of predicted amenities (using all controls of Table 2, column 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 (column 1).

ONLINE APPENDIX—not for publication

A Additional figures and tables

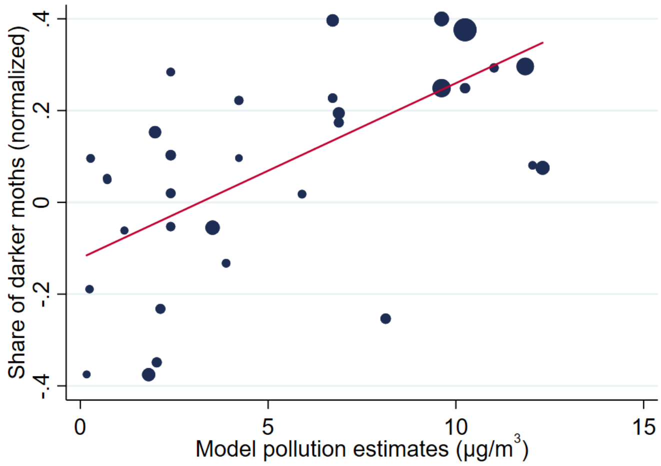
Figure A1. Topography and historical air pollution—the example of Oldham.



(a) Industrial pollution. (b) Residential pollution.

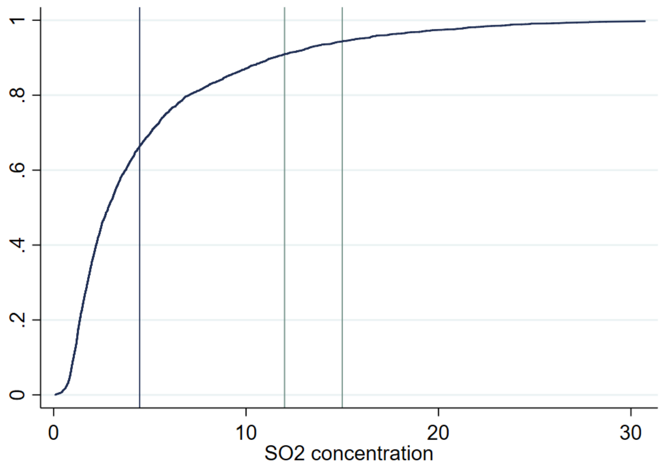
Sources: ADMS 5. These maps show elevation (from green to brown) and the level lines for industrial and residential pollution in Oldham.

Figure A2. Share of darker moths (*carbonaria*) and air pollution—external validity.



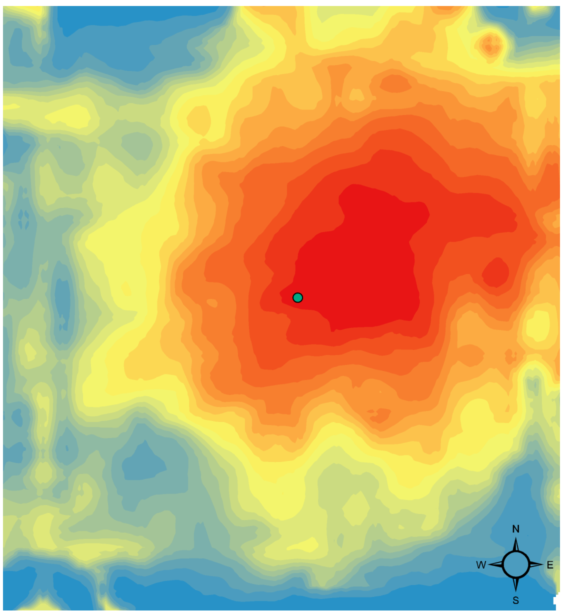
Sources: This Figure shows the relationship between the darker melanic form in species of moths and our measure of $SO2$ concentration in $\mu g/m^3$. We use 54 geolocated surveys of moths, collected between 1965 and 1975 across various sites around English cities and with more than 30 observations (Cook, 2018). We construct the share of the darker form (*carbonaria*). As we condition on the species (mostly *Biston betularia* and *Odontoptera bidentata*), the y-axis represents the residual share of the darker form within species. The correlation between the two measures is 0.66, and is weighted by the number of moths identified at each survey site.

Figure A3. Cumulative of pollution in our baseline sample of 5,538 LSOAs and National Ambient Air Quality Standards (12-15 $\mu g/m^3$).

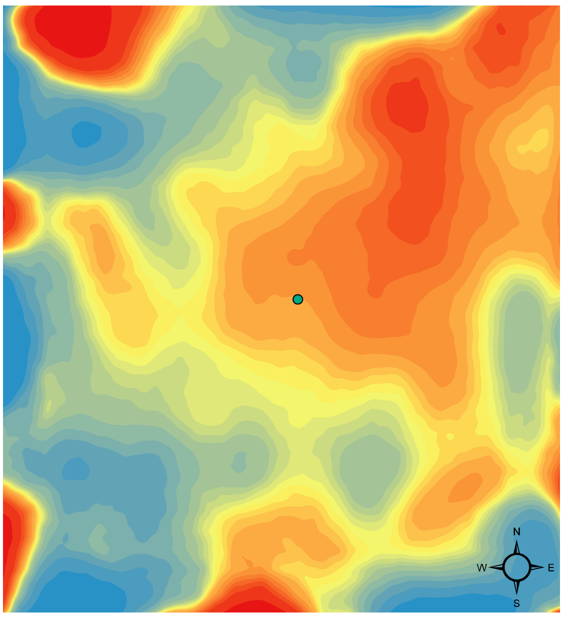


Sources: This Figure represents the cumulative distribution of SO2 concentration as predicted by the distribution of pollution sources, the emission intensity and air pollutant dispersion. See Section 1 and Appendix Section C for additional details.

Figure A4. Pollution and shares of low-skilled workers in the average city—robustness check using town halls in order to calibrate grids across cities.



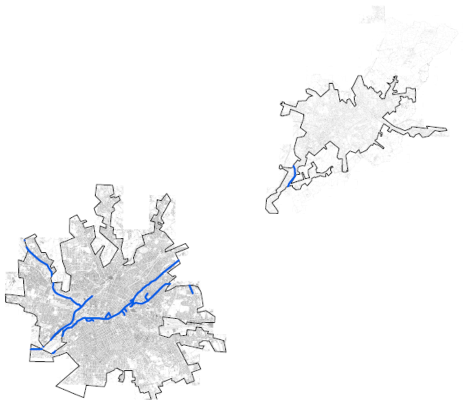
(a) Pollution.



(b) Share of low-skilled workers.

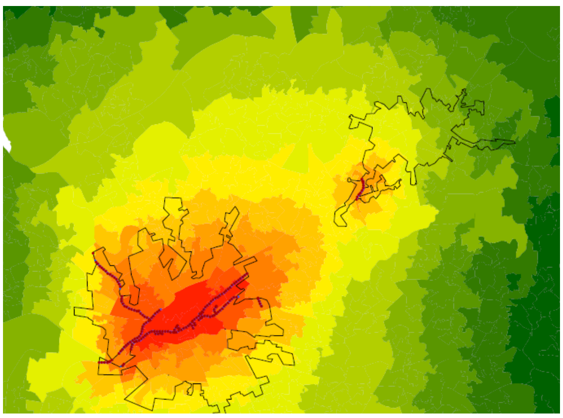
Notes: The left panel of this Figure displays the average gradient of pollution in 1880–1900 across cities. The right panel displays the average gradient of neighborhood composition—as captured by the share of low-skilled workers in 1881—across cities. To construct this Figure, we define a grid of equally-spaced points within 1.5 kilometers of the town hall of each city (indicated with a green circle); we overlay the grids across cities and consider the unweighted average. In order to create a continuous measure, we interpolate using a Gaussian Kernel. The 20 level lines are quantiles of pollution (left panel) and share of low-skilled workers (right panel).

Figure A5. The IV empirical approach with hypothetical chimneys located along old waterways—illustration in Manchester and Oldham.

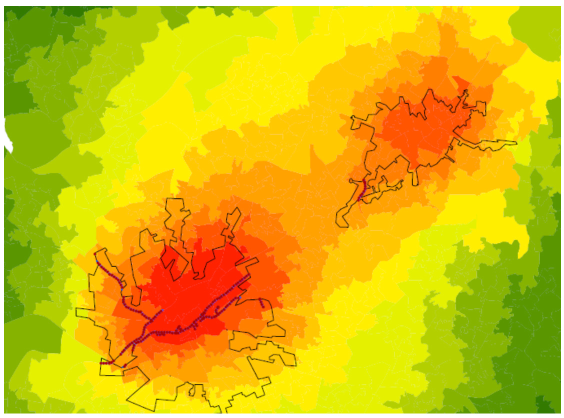


(a) City map and waterways

(b) Hypothetical chimneys along waterways.



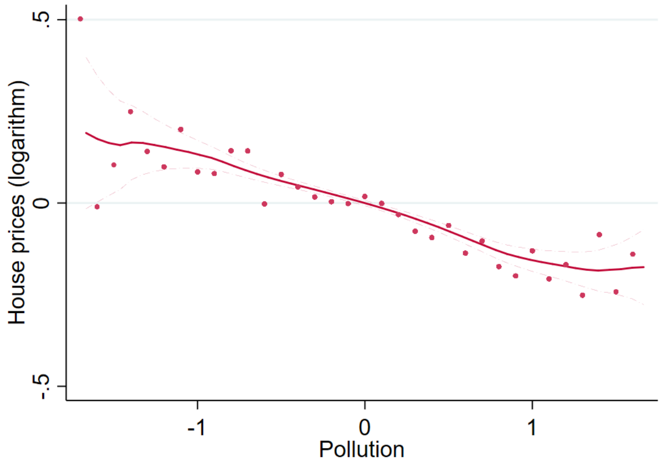
(c) Predicted pollution instrument.



(d) Simulated historical pollution.

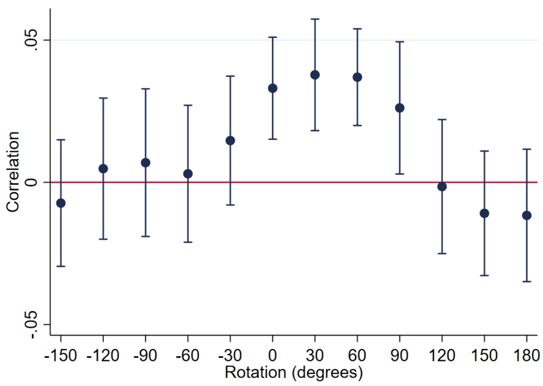
Sources: Authors' calculations using Ordnance Survey Maps—25 inch to the mile, 1842–1952 and the ADMS 5 Air Pollution Model. Chimneys are indicated with a red dot, and 1827 natural waterways with blue lines.

Figure A6. House transaction prices between 2000 and 2011 (y-axis) and pollution (x-axis) across neighborhoods.

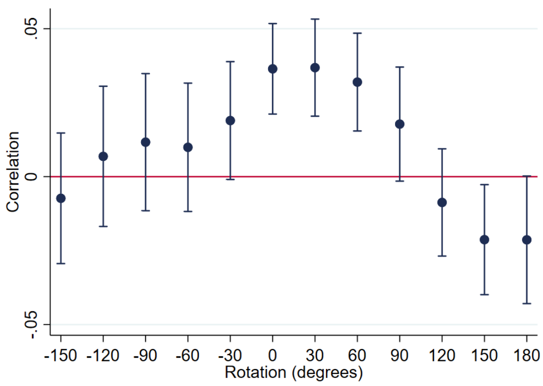


Notes: This Figure represents the relationship between the average (log) transaction prices between 2000 and 2011 and our (standardized) measure of past pollution. We consider the residuals of all measures once cleaned by city fixed effects, geographic and topographic controls. We group neighborhoods, create 100 bins of neighborhoods with similar past pollution and represent the average house prices within a pollution-bin. The lines are locally weighted regressions on all observations.

Figure A7. Rotating wind patterns in 1991 and 2011.



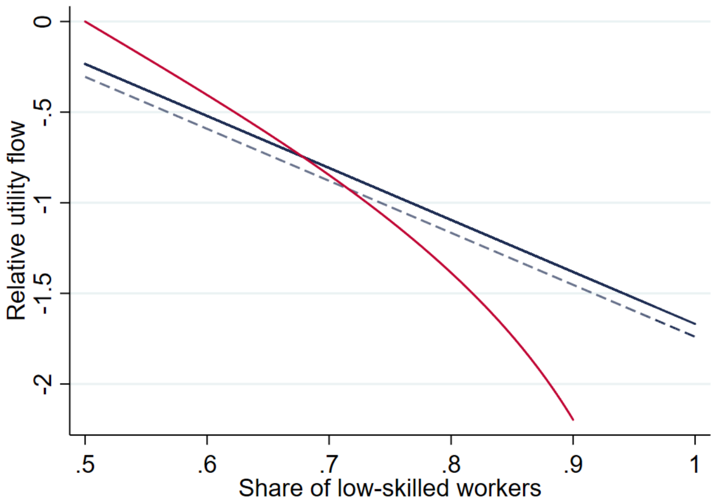
(a) 1991



(b) 2011

Notes: These Figures represent the conditional correlations between the shares of low-skilled workers (in 1991 and in 2011) and counterfactual measures of past pollution rotated in steps of 30 degrees. Each dot represents the estimate in a specification including the controls reported in Table 2, Column 6, and the measure of *symmetric pollution* capturing proximity to the pollution source (see Appendix Table A7). Standard errors are clustered at the parish-level, and the lines represent 5% confidence intervals.

Figure A8. Illustration of the steady-state equilibrium as induced by our estimates of neighborhood demand.

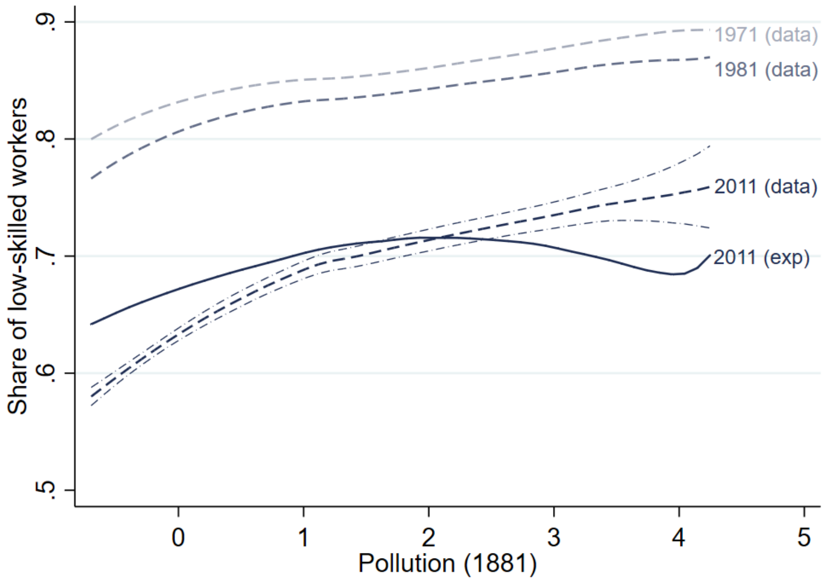


Notes: This Figure represents the steady-state share of low-skilled workers, $s_{j,c}$, at the equilibrium. More precisely, it represents the equation,

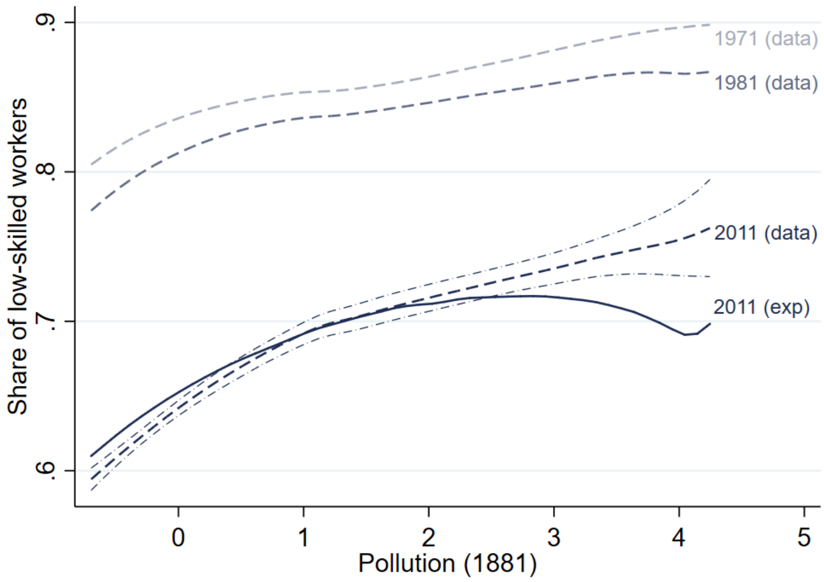
$$(1-\theta\beta)\ln\left(\frac{1-s_{j,c}}{s_{j,c}}\right) = h(a_{j,c}, s_{j,c}) + \nu_c.$$

evaluated using the estimates of Table 7, column 1. The red line represents the left-hand side of the equation while the blue line represents the right-hand side. The dashed line shows the effect of an off-equilibrium increase in the share of low-skilled workers (one standard deviation or 7 percentage points) on the relative valuation (a shift in demand) and the subsequent negative adjustment in the share of high-skilled workers (of about 3 percentage points).

Figure A9. Simulated dynamics (1971–2011) using (a) different preferences, (b) different relocation frictions.



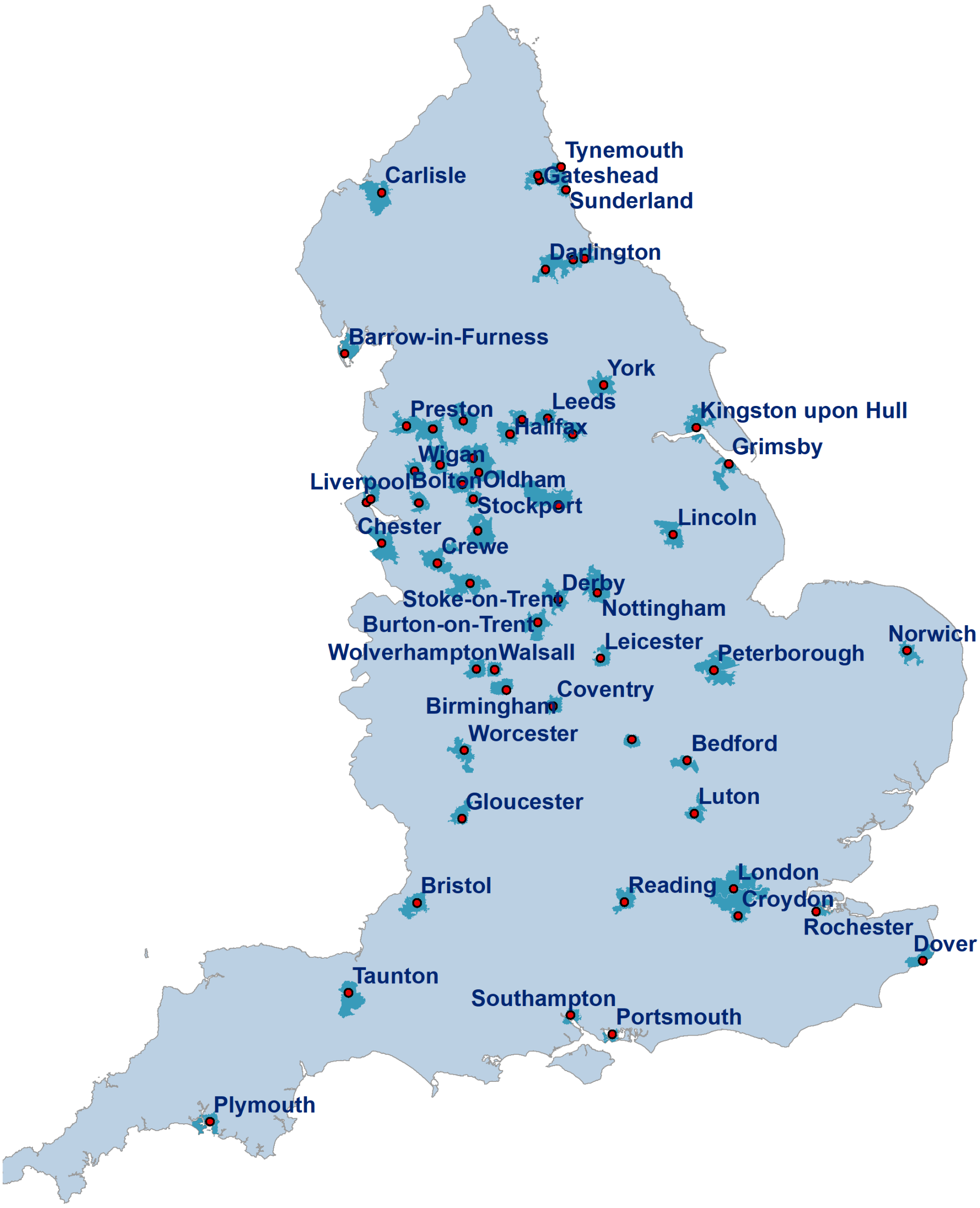
(a) No neighborhood effects



(b) Low relocation rigidities

Notes: This Figure shows the average shares of low-skilled workers as a function of past pollution, using a local polynomial smoothing, in two counterfactual experiments. Panel (a) (resp. b) displays counterfactual transitional dynamics computed with preferences that are set to be orthogonal to current neighborhood composition (resp. lower relocation rigidities, i.e., $\theta = 0.25$). Dashed lines (1971, 1981, 2011 from lighter to darker blue) represent observed shares of low-skilled workers; plain lines represent simulated shares in 2011.

Figure A10. Map of the sample of 5,538 LSOAs across England.



Notes: The baseline sample covers 70 metropolitan areas, 142 "cities" or local administrations, 542 parishes and 5,538 LSOAs.

Table A1. Average coal use per industry, and estimated average coal use per chimney.

Industrial categories	Average coal use C_i $m.tons/year$	Average weight E_i
Breweries	19.4	0.36
Brick factories	48.9	1.05
Chemical factories	40.1	0.84
Food processing	12.0	0.77
Foundries	43.7	1.00
Mining	28.9	3.55
Paper production	9.7	2.47
Shipbuilding	6.1	1.02
Tanneries	12.1	0.17
Textile production	10.1	0.47
Wood processing	5.4	0.10
Other manufactures	-	0.10

Source: Hanlon (2019) and the 1907 Census of Production. Notes: Average coal use per worker C_i is reported in tons per year and the estimated coal use per chimney E_i is normalized such as to be equal to 1 for the category "Foundries". The measure E_i is set to the minimum average value for chimneys classified as "Other manufactures".

Table A2. Pollution and shares of low-skilled workers in 1881 and 2011—the role of covariates.

Share of low-skilled workers	1881	2011
Pollution	.0337	.0354
	(.0069)	(.0057)
Employment (thousand, 1881)	-.0007	.0007
	(.0005)	(.0004)
Share low-skilled (1817)	.2455	-.0111
	(.0888)	(.0887)
Share farmers (1817)	-.0783	-.1614
	(.0954)	(.0986)
Property tax (1815)	.0091	-.0111
	(.0073)	(.0057)
Maximum elevation	-.0001	-.0006
	(.0004)	(.0002)
Minimum elevation	-.0010	.0003
	(.0004)	(.0004)
Average elevation	.0002	-.0005
	(.0007)	(.0004)
Distance waterways (inverse)	.0705	.0351
	(.0170)	(.0313)
Distance hall (inverse)	-.0083	.0250
	(.0283)	(.0202)
Distance parks (inverse)	-.0306	-.0432
	(.0120)	(.0104)
Share area (city)	.0202	-.0285
	(.0126)	(.0135)
Area	-.0021	-.0002
	(.0009)	(.0008)
Distance heavy industry (inverse)	.0017	-.0219
	(.0372)	(.0631)
Distance light industry (inverse)	.3454	.1378
	(.1448)	(.1163)
Longitude	.0021	.0064
	(.0022)	(.0023)
Latitude	.0030	.0023
	(.0023)	(.0020)
Observations	5,538	5,538

Standard errors are reported between parentheses and are clustered at the parish-level. Standardized effects are reported between square brackets. Each cell is the result of a separate regression. The unit of observation is a Lower Super Output Area.

Table A3. Pollution and other outcomes in 1881.

VARIABLES	Low-skilled, all	Migrants: All	Migrants: England/Wales	Migrants: Commonwealth
Pollution	.0242	.0314	.0282	.0033
	(.0051)	(.0072)	(.0071)	(.0019)
	[.1192]	[.1103]	[.1101]	[.0433]
Observations	5,538	4,312	4,312	4,312
Fixed effects (city)	Yes	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes	Yes

Standard errors are reported between parentheses and are clustered at the parish-level. Standardized effects are reported between square brackets. Each cell is the result of a separate regression. The unit of observation is a Lower Super Output Area. *Low-skilled, all* is the share of all workers between 25 and 55 years old that are employed in low-skilled occupations. Migrants are defined as individuals between 25 and 55 years old who are born in a different county. The set of extended controls include all controls of column 6 in Table 2.

B Appendix on the dynamic model of residential sorting

In this section, we describe the derivation of the neighborhood choice equation (1) from the household maximization problem (Section B.1); we provide the detailed derivation of the dynamic demand equation using equilibrium and land market clearing conditions (Section B.2); we discuss extensions of the baseline model (migration across cities, many household types, elastic land supply, see Section B.3); we provide a characterization of the steady-state residential sorting and the stability of the dynamic system induced by the dynamic demand equation (Section B.4); and we discuss the local identification of Equation (S3) (Section B.5).

B.1 Neighborhood choice

This section derives Equation (1) from the household maximization problem. A household without the opportunity to relocate in period t does not take any decision. A household subject to a relocation choice maximizes the expected stream of utility, taking as given the path of future (and possibly stochastic) amenities $\{a_{j,t}\}_{j,t}$. The household maximizes:

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

As re-optimization draws and preference shocks are independent (across time, across households, and between them), $E_t V_{t+1}$ and $E_t U_{j,t+1}$ are common to all households of a given type—as is $u_{j,t}$. As the preference draw $\varepsilon_{j,t}$ is distributed along a Type 1 Extreme-Value distribution, the fraction of households opting for neighborhood j among the ones that were given such choice, $n^*_{j,t}$, follows:

$$n^*_{j,t} = \frac{e^{u_{j,t}+\beta\theta E_t U_{j,t+1}+\beta(1-\theta)E_t V_{t+1}}}{\sum_j e^{u_{j,t}+\beta\theta E_t U_{j,t+1}+\beta(1-\theta)E_t V_{t+1}}}.$$

Equation (1) can then be obtained in three steps. In the first step, we iterate forward to find an expression for the expected value of residing in neighborhood j in period $t+1$, $E_t U_{j,t+1}$. This value verifies,

$$E_t U_{j,t+1} = E_t \left[u_{j,t+1} + \beta\theta E_{t+1} U_{i,j,t+2} + \beta(1-\theta) E_{t+1} V_{t+2}\right]$$

given that preference shocks are independent across periods, and that the probability to be able to relocate in each period is $(1-\theta)$. Using the law of iterated expectations, we obtain:

$$E_t U_{j,t+1} = E_t u_{j,t+1} + \beta (1-\theta) E_t V_{t+2} + \beta\theta E_t U_{j,t+2}$$

$$E_t U_{j,t+1} = \sum_{\tau=1}^{\infty} (\beta\theta)^{\tau-1} E_t \left[u_{j,t+\tau} + \beta (1-\theta) E_t V_{t+\tau+1}\right]$$

In a second step, we substitute $E_t U_{j,t+1}$ by the previous expression,

$$n^*_{j,t} = \frac{e^{u_{j,t}+\beta(1-\theta)E_t V_{t+1}+\sum_{\tau=1}^{\infty}(\beta\theta)^{\tau} E_t[u_{j,t+\tau}+\beta(1-\theta)E_t V_{t+\tau+1}]}}{\sum_j e^{u_{j,t}+\beta(1-\theta)E_t V_{t+1}+\sum_{\tau=1}^{\infty}(\beta\theta)^{\tau} E_t[u_{j,t+\tau}+\beta(1-\theta)E_t V_{t+\tau+1}]}}.$$

where we can see that neighborhood choice is a function of: (a) the expected stream of period utility weighted by the probability to be unable to move away from the neighborhood; (b) the values of re-optimizing (independent across neighborhoods).

In a third step, we simplify by these values of re-optimizing and divide numerator and denominator by $e^{\sum_{\tau=0}^{\infty}(\beta\theta)^{\tau}\beta(1-\theta)E_t V_{t+\tau+1}}$:

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

Two assumptions allow us to derive a dynamic discrete choice equation with a multinomial structure: re-optimization draws and preference shocks are idiosyncratic (independent across time, across households, and between them), preference shocks are distributed along a Type 1 Extreme-Value distribution.

B.2 Equilibrium, demand for neighborhoods and empirical specification

This section provides the detailed derivation of the dynamic demand equation using the equilibrium and land market clearing conditions.

Equilibrium and land market clearing We normalize total population to $N = 1$ and consider two types of households $o \in \{H, L\}$ in fixed aggregate proportions $N^H = \eta \in (0,1)$ and $N^L = (1-\eta)$. Let $N^o_{j,t}$ denote the number of o-type households in j at t, so that $n^o_{j,t} = N^o_{j,t}/N^o$ is the share of o-type households in neighborhood j at time t. At each period t, the number of o-type households is composed of a share θ of non-movers from the previous period, and a share $1-\theta$ of households having just opted for neighborhood j. Consequently, the law of motion requires that $n^{*,o}_{j,t} = \frac{n^o_{j,t}-\theta n^o_{j,t-1}}{1-\theta}$, and the respective demand by each type for neighborhood j

is given by Equation (1),

$$\begin{cases} \ln\left(\dfrac{n^H_{j,t}-\theta n^H_{j,t-1}}{1-\theta}\right) = \sum_{\tau=0}^{\infty}(\beta\theta)^{\tau}\, E_t u^H_{j,t+\tau} - \mu^H_t \\ \ln\left(\dfrac{n^L_{j,t}-\theta n^L_{j,t-1}}{1-\theta}\right) = \sum_{\tau=0}^{\infty}(\beta\theta)^{\tau}\, E_t u^L_{j,t+\tau} - \mu^L_t \end{cases}$$

where μ^o_t are type-specific aggregate present discounted values.

The land market clearing condition for each neighborhood j, $N^H_{j,t} + N^L_{j,t} = N_j$, can thus be written, using $n^o_{j,t} = N^o_{j,t}/N^o$, as,

$$\eta n^H_{j,t} + (1-\eta) n^L_{j,t} = n_j,$$

where n_j is the fixed land supply in neighborhood j.

Relative demand for neighborhoods We use these equilibrium conditions in periods t and $t-1$ in order to express the whole problem as a function of the share of type-L households in neighborhood j and period t, $s_{j,t} = N^L_{j,t}/N_j = (1-\eta)n^L_{j,t}/n_j$. Note that, since $n^L_{j,t} = \frac{n_j s_{j,t}}{1-\eta}$, we can re-write the law of motion for type-L households in terms of shares,

$$s_{j,t} = \theta s_{j,t-1} + (1-\theta)\left(\frac{1-\eta}{n_j}\right) n^{*,L}_{j,t},$$

from which we have an expression for $n^{*,L}_{j,t}$ in terms of $s_{j,t}$ and $s_{j,t-1}$. The land market clearing condition together with the law of motion for each type means that, in each neighborhood j and at each period t,

$$\eta\left[\theta n^H_{j,t-1} + (1-\theta) n^{*,H}_{j,t}\right] + (1-\eta)\left[\theta n^L_{j,t-1} + (1-\theta) n^{*,L}_{j,t}\right] = n_j,$$

which, using $\eta n^H_{j,t-1} + (1-\eta) n^L_{j,t-1} = n_j$, reduces to,

$$\eta n^{*,H}_{j,t} + (1-\eta) n^{*,L}_{j,t} = n_j.$$

The land market clearing conditions holding in period $t-1$ and period t impose that an equivalent condition must hold for re-optimizing households only—non-movers occupy $1-\theta$ of the land in every neighborhood. The resulting demand equations in terms of shares are,

$$\begin{cases} \ln\left(\frac{n_j}{\eta}\left(1-\frac{s_{j,t}-\theta s_{j,t-1}}{1-\theta}\right)\right) = \sum_{\tau=0}^{\infty}(\beta\theta)^{\tau} E_t u^H_{j,t+\tau} - \mu^H_t \\ \ln\left(\frac{n_j}{1-\eta}\frac{s_{j,t}-\theta s_{j,t-1}}{1-\theta}\right) = \sum_{\tau=0}^{\infty}(\beta\theta)^{\tau} E_t u^L_{j,t+\tau} - \mu^L_t \end{cases}$$

Subtracting the two demand equations yields the relative demand for neighborhood j from type-H households,

$$F\left(\frac{s_{j,t}-\theta s_{j,t-1}}{1-\theta}\right) = \sum_{\tau=0}^{\infty}(\beta\theta)^{\tau} E_t\left(u^H_{j,t+\tau}-u^L_{j,t+\tau}\right)+\mu_t,$$

where $F(x)=\ln(1-x)-\ln(x)$, and $\mu_t=\mu^L_t-\mu^H_t+\ln(\eta/(1-\eta))$ captures the relative welfare of each type in period t. Using the law of iterated expectations, the relative demand verifies the following forward-looking recursive equation:

$$F\left(\frac{s_{j,t}-\theta s_{j,t-1}}{1-\theta}\right)-\mu_t = u^H_{j,t}-u^L_{j,t}+\beta\theta E_t\left[F\left(\frac{s_{j,t+1}-\theta s_{j,t}}{1-\theta}\right)-\mu_{t+1}\right].$$

The dynamic demand Equation (2) for neighborhood j immediately derives from the previous expression.

Empirical specification We now connect the previous equation to our empirical specification (S3). Let $u^o_{j,c,t}$, the utility of a type-o household living in neighborhood j of city c, be,

$$u^o_{j,c,t} = \left[a_{j,c}+\chi_{j,c,t}+g(s_{j,c,t})\right](y^o-r_{j,c,t}),$$

where the perceived neighborhood amenity $a_{j,c}+\chi_{j,c,t}+g(s_{j,c,t})$ is independent of the household type o. The right-hand side of Equation (2) becomes:

$$\left[a_{j,c}+\chi_{j,c,t}+g(s_{j,c,t})\right]\left(y^H-r_{j,c,t}-(y^L-r_{j,c,t})\right)+\nu_{c,t},$$

and we obtain,

$$F\left(\frac{s_{j,c,t}-\theta s_{j,c,t-1}}{1-\theta}\right)-\beta\theta E_t F\left(\frac{s_{j,c,t+1}-\theta s_{j,c,t}}{1-\theta}\right) = \left[a_{j,c}+\chi_{j,c,t}+g(s_{j,c,t})\right]\left(y^H-y^L\right)+\nu_{c,t}.$$

Letting $b_{j,c}=a_{j,c}\left(y^H-y^L\right)$ and using a log-linear approximation for function g, i.e., $g(s_{j,c,t})\propto s^{\gamma}_{j,c,t}$, we have:

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

where the noise $\zeta_{j,c,t}$ is defined as,

$$\zeta_{j,c,t} = \chi_{j,c,t}\left(y^H - y^L\right) + \beta\theta F\left(\frac{s_{j,c,t+1} - \theta s_{j,c,t}}{1-\theta}\right) - \beta\theta E_t F\left(\frac{s_{j,c,t+1} - \theta s_{j,c,t}}{1-\theta}\right),$$

thus satisfying $E_{t-1}\left[\zeta_{j,c,t}\right] = 0$.[40] Note that we can, in principle, allow the parameters (α_0, α_1) to be a function of *observable* neighborhood characteristics.

The error term is a combination of two terms: (i) a neighborhood amenity shock which is observable to households but not to the econometrician and (ii) a future realization shock, as we use actual realizations of future neighborhood compositions rather than their expectations at period t.

B.3 Extensions of the baseline model

In this section, we discuss two natural extensions of the baseline model. In a first step, we allow for the shares of low-skilled workers to fluctuate at the city level, as induced by (endogenous) migration across cities. In a second step, we relax the assumption that there are only two household types and that land/housing supply is constant.

Migration across cities We now describe how the baseline model could accommodate an endogenous composition of cities, without modifying the empirical specification. Consider the following extension of the model: there is a unit mass of infinitely-lived households, each of measure zero; time is discrete; there is a discrete number of cities C, each city c with a discrete number of neighborhoods J_c; land supply is inelastic and constant over time; preferences are similar as in the baseline model.

The timing of actions is as follows. At the beginning of each period t, there is a household-specific idiosyncratic draw: with probability $1-\theta$, the household can freely relocate to any city and any neighborhood within the period. The possible move is then staggered. First, the independent, idiosyncratic preference shocks for cities are revealed. Households choose to migrate to a city. Second, idiosyncratic preference shocks for neighborhoods are revealed and households relocate to a neighborhood—the latter shocks are distributed across households along a Type 1 Extreme-Value distribution. This structure gives rise to a nested logit structure (à la Monras, 2020) where the decision to choose a city is independent from the decision to choose a specific neighborhood within cities.

[40] We assume that agents know the distribution of amenity shocks and form expectations accordingly.

The possibility to migrate across cities modifies our household maximization as follows. Let $V_{c,t}$ denote the expected value of choosing city c before preferences for neighborhoods are revealed and let $\chi_{c,t}$ denote the city preference draws. The household chooses cities by maximizing:

$$V_t = \max_c \{V_{c,t} + \chi_{c,t}\}$$

Once households have chosen a certain city c, they solve:

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

where $E_t V_{t+1}$, $u_{j,c,t}$ and $E_t U_{j,c,t+1}$ are common to all households of type j. Following the argument developed in Appendix B.1, we obtain that the fraction of household opting for neighborhood j, conditional on having opted for city c, is:

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

In equilibrium, land markets must clear such that population remains constant—the number of immigrants should equal the number of emigrants. Consequently, the share of high-skilled workers in period t, $\eta_{c,t}$, verifies:

$$\eta_{c,t} = \theta\eta_{c,t-1} + (1-\theta)\eta^*_{c,t}$$

where $\eta^*_{c,t}$ is the share of high-skilled workers among movers. The law of motion for the share of type-L households in neighborhood j verifies,

$$s_{j,c,t} = \theta s_{j,c,t-1} + (1-\theta)\left(\frac{1-\eta^*_{c,t}}{n_{c,j}}\right) n^{*,L}_{c,j,t},$$

because the overall fraction of low-skilled workers among movers, $1-\eta^*_t$, is now endogenous. We can follow the same steps as in Appendix B.2, where we see that $\eta^*_{c,t}$ is absorbed in a city $\times$ year fixed-effect $\mu_{c,t}$:

$$F\left(\frac{s_{j,c,t} - \theta s_{j,c,t-1}}{1-\theta}\right) = \sum_{\tau=0}^{\infty} (\beta\theta)^\tau E_t\left(u^H_{j,c,t+\tau} - u^L_{j,c,t+\tau}\right) + \mu_{c,t},$$

where $F(x) = \ln(1-x) - \ln(x)$, and $\mu_t = \mu^L_t - \mu^H_t + \ln\left(\eta^*_{c,t}/(1-\eta^*_{c,t})\right)$ captures the relative welfare of each type in period t.

Consequently, the endogenous changes in aggregate composition at the city level,

$\eta^*_{c,t}$, are fully absorbed by city $\times$ year fixed effects, and our empirical specification remains unchanged, irrespective of the drivers of allocation across cities.

Many household types and elastic land supply Consider I types of households populating a city, let η_i denote the total share of households i within the city. The distribution of the idiosyncratic preference shocks implies that the share of type-i households opting for neighborhood j verifies,

$$x^*_{i,j,t} = \frac{e^{u_{i,j,t}+\beta\theta E_t U_{i,j,t+1}+\beta(1-\theta)V_{i,t+1}}}{\sum_j e^{u_{i,j,t}+\beta\theta E_t U_{i,j,t+1}+\beta(1-\theta)V_{i,t+1}}}.$$

Consequently,

$$\ln\left(x^*_{i,j,t}\right) = u_{i,j,t} + \beta\theta E_t U_{i,j,t+1} + \mu_{i,t},$$

where $\mu_{i,t}$, capturing $V_{i,t+1}$ and the denominator, is independent from neighborhood j. Let $u_{i,j,t}$, the utility of a type-o household living in neighborhood j of city c, be,

$$u_{i,j,t} = \left[a_j + \chi_{j,t} + g(\mathbf{x}_{j,t})\right]\left(y_i - r_{j,t}\right),$$

where $r_{j,t}$, the rental price in neighborhood j, is independent of household type i and $\mathbf{x}_{j,t} = \{x_{i,j,t}\}_i$ capture the share of households living in neighborhood j at period t. For each household type i, we have that:

$$\ln\left(\frac{x_{i,j,t} - x_{i,j,t-1}}{1-\theta}\right) - \beta\theta E_t \ln\left(\frac{x_{i,j,t+1} - x_{i,j,t}}{1-\theta}\right) = \left[a_j + \chi_{j,t} + g(\mathbf{x}_{j,t})\right]\left(y_i - r_{j,t}\right) + w_{i,t} \quad (3)$$

where $w_{i,t} = \mu_{i,t} - \beta\theta E_t\mu_{i,t+1} + \ln(\frac{\eta_i}{n_{j,t}}) - \beta\theta E_t \ln(\frac{\eta_i}{n_{j,t+1}})$. The previous set of equations characterizes the demand for neighborhood j at period t.

We assume that land supply is elastic and verifies,

$$n_{j,t}/n_j = \left(r_{j,t}/r\right)^{\nu},$$

with $\nu > 0$. Land market clearing imposes that:

$$\sum_i \mu_i x_{i,j,t} = n_{j,t}/n_j \quad (4)$$

such that land supply writes as,

$$\ln\left(\sum_i \mu_i x_{i,j,t}\right) = \nu \ln\left(r_{j,t}/r\right). \quad (5)$$

The three sets of equations (3)—land demand for type i and neighborhood j, (5)—land supply for neighborhood j, and equilibrium conditions (4) perfectly characterize neighborhood sorting and its evolution over time. The estimation of this system however requires observing rental prices, or the housing stock, and their evolution over time.

B.4 Steady-state and dynamic stability

In this section, we establish the properties of the recursive process described by Equation (2). First, the system admits at least one stationary point which verifies, for each neighborhood j:

$$(1-\theta\beta)\ln\left(\frac{1-s_j}{s_j}\right) = h\left(a_j, s_j\right) + \nu. \tag{6}$$

The left-hand side of the equation describes a strictly decreasing function with limits $-\infty$ and $+\infty$ in 0 and 1. The right-hand side of the equation describes an empirical relationship which needs to be estimated. One can conjecture that h is monotone and decreasing in s_j, with defined limits in 0 and 1 but, again, that needs not be the case. For this reason, there may exist several stationary points (an uneven number). In our empirical application, the left-hand side equation will always be steeper than $h(\cdot)$: there will exist a single share s_j which verifies the steady-state equation for each neighborhood. Each of these shares is positively related to the average (relative) welfare of type-H households, as captured by $-\nu$. The equilibrium condition $\sum_j n_{j,t}^H = 1$ thus ensures that there is a unique allocation $\{s_j\}_{j=1,\ldots,J}$, or equivalently $\left\{n_{j,t}^H\right\}_{j=1,\ldots,J}$, verifying the system of Equations (6).

Second, Equation (2) describes a non-linear recursive sequence of order 2 for each neighborhood j. We linearize this equation around the steady-state in order to study its global asymptotic properties:

$$E_t\left[s_{j,t+1} - s_j\right] + b\left(s_{j,t} - s_j\right) + c\left(s_{j,t-1} - s_j\right) = 0,$$

where $b = -\frac{1}{\theta\beta}\left[1 - \beta\theta^2 + \frac{(1-\theta)h'(s_j)s_j(1-s_j)}{(1-\theta\beta)}\right]$ and $c = 1/\beta$. The characteristic polynomial associated with the sequence is $X^2 + bX + c$ which admits two real roots, (r_1, r_2),

$$\begin{cases} r_1 = \frac{-b-\sqrt{b^2-4c}}{2} \\ r_2 = \frac{-b+\sqrt{b^2-4c}}{2} \end{cases}$$

The (deterministic) solutions to the previous sequence are of the general form $k_1(r_1)^t + k_2(r_2)^t$. Under some estimates of $(\beta, \theta, h'(s_j))$, the roots will be ordered as follows, $0 < r_1 < 1 < r_2$. In such case, the equivalent of a "no-Ponzi condition", anchoring future expectations in the absence of shocks, imposes that $k_2 = 0$ such that we have:

$$s_{j,t+1} - s_j = r_1 (s_{j,t} - s_j) + \chi_{t+1},$$

with $E_t [\chi_{t+1}] = 0$. The initial share of households, $s_{j,0}$, would then be sufficient to fully characterize the expected dynamics of the system.

B.5 Local identification

In this section, we discuss more formally the identification of Equation (S3), and we provide some intuition behind the separate identification of the discount factor, β, the relocation rigidities, θ, and preferences for neighborhood quality.

Consider the function $g(\mathbf{s}, \theta, \beta, \alpha_0, \alpha_1, \gamma)$ defined by:[41]

$$g(s, \theta, \beta, \alpha_0, \alpha_1) = F\left(\frac{s_{j,c,t} - \theta s_{j,c,t-1}}{1-\theta}\right) - \beta\theta E_t F\left(\frac{s_{j,c,t+1} - \theta s_{j,c,t}}{1-\theta}\right) - \alpha_0 - \alpha_1 s^{\gamma}_{j,c,t}$$

The local identification of the vector $(\theta, \beta, \alpha_0, \alpha_1, \gamma)$ requires that the columns of matrix $WE[\nabla g(s, \theta, \beta, \alpha_0, \alpha_1, \gamma)]$ are independent. We calculate below the partial derivatives with respect to each parameter,

$$\begin{aligned}\frac{\partial g(s,\theta,\beta,\alpha_0,\alpha_1)}{\partial\theta} = \; & F'\left(\frac{s_{j,c,t}-\theta s_{j,c,t-1}}{1-\theta}\right)\frac{s_{j,c,t}-s_{j,c,t-1}}{(1-\theta)^2} - \beta\theta E_t F'\left(\frac{s_{j,c,t+1}-\theta s_{j,c,t}}{1-\theta}\right)\frac{s_{j,c,t+1}-s_{j,c,t}}{(1-\theta)^2} \\ & -\beta E_t F\left(\frac{s_{j,c,t+1}-\theta s_{j,c,t}}{1-\theta}\right)\end{aligned}$$

where $F'(x) = -\frac{1}{1-x} - \frac{1}{x}$,

$$\frac{\partial g(s, \theta, \beta, \alpha_0, \alpha_1)}{\partial\beta} = -\theta E_t F\left(\frac{s_{j,c,t+1} - \theta s_{j,c,t}}{1-\theta}\right)$$

$$\frac{\partial g(s, \theta, \beta, \alpha_0, \alpha_1)}{\partial\alpha_0} = 1$$

$$\frac{\partial g(s, \theta, \beta, \alpha_0, \alpha_1)}{\partial\alpha_1} = s^{\gamma}_{j,c,t}$$

$$\frac{\partial g(s, \theta, \beta, \alpha_0, \alpha_1)}{\partial\gamma} = \alpha_1 s^{\gamma}_{j,c,t} \ln(s_{j,c,t})$$

As long as $0 < \theta < 1$ (i.e., there are some relocation rigidities), local identification

[41] For the sake of exposition, we ignore the estimation of city × year fixed-effects and neighborhood fixed effects.

requires the following rows to be independent within our estimation sample:

$$\begin{pmatrix} F'\left(\frac{s_{j,c,t}-\theta s_{j,c,t-1}}{1-\theta}\right)\frac{s_{j,c,t}-s_{j,c,t-1}}{(1-\theta)^2} - \beta\theta E_t F'\left(\frac{s_{j,c,t+1}-\theta s_{j,c,t}}{1-\theta}\right)\frac{s_{j,c,t+1}-s_{j,c,t}}{(1-\theta)^2} \\ E_t F\left(\frac{s_{j,c,t+1}-\theta s_{j,c,t}}{1-\theta}\right) \\ s_{j,c,t} \\ \alpha_1 s_{j,c,t}^{\gamma}\ln(s_{j,c,t}) \\ 1 \end{pmatrix}$$

One can see from the previous expressions that:

- the identification of preferences for neighborhood quality, $(\alpha_0, \alpha_1, \gamma)$, relies on variation in neighborhood composition, $s_{j,c,t}$;
- the identification of the discount factor, β, relies on variation in the "slope" of residential dynamics, $\frac{s_{j,c,t+1}-\theta s_{j,c,t}}{1-\theta}$, coupled with the specific functional form F, as induced by the Extreme Value distribution of preferences;
- the identification of the discount factor, θ, mostly relies on variation in the "curvature" of residential dynamics coupled with the specific functional form, F.

In order to better understand the latter argument, let $\beta \approx 1$ and consider a first-order approximation around a steady-state value $\bar{s}$. The first row of the previous vector becomes,

$$\frac{F'(\bar{s})}{(1-\theta)^2}\left(s_{j,c,t} - s_{j,c,t-1} - \theta E_t[s_{j,c,t+1} - s_{j,c,t}]\right).$$

The main empirical moments that underlie the estimation of Equation (S3) are thus the level of deprivation, its growth rate and its possible acceleration or slowdown over time. Figures 10 and 11 show that historical pollution strongly predicts the different moments of the dynamics in $s_{j,c,t-1}$, i.e., its initial level in 1971, its variation from one period to the next and possible non-linear effects in such variation.

C Data sources

In this section, we describe the data construction of pollution sources and amenities in 1880–1900; census data in 1817 and 1881; bomb damage in 1940–1945; steam engines in 1700–1800 and waterways; and more recent data covering the period 1971–2016.

Pollution sources and amenities in 1880–1900 We rely on scans of map tiles of the collection "Ordnance Survey maps—25 inch to the mile (1842–1952)". These maps are drawn differently across counties, covering different periods and different waves. For instance, Bedfordshire is covered by four waves, 1878–1883, 1899–1901, 1922–1924 and 1937–1940. All counties were supposed to be revised every twenty years on average, but rural maps were infrequently revised.

We consider the 70 largest industrial centers at the beginning of the nineteenth century, and select the nearest wave to the year 1890.[42] While the level of precision is unmatched (the positions of free-standing trees are reported), it comes at the expense of reproduction quality. The printing process involved zinc plates, and the resulting quality is not fitted for direct recognition processes or automatic map digitization.

In order to alleviate this issue, we design a recognition process which works as follows. In a first step, for each tile, we mark interesting landmarks with a recognizable sign (e.g., a red cross X) and an associated identifier (e.g., 00001), and we report information about the landmarks in a separate excel file (with information about the identifier, the type of landmarks and the name). The following landmarks are marked and digitized: Chimneys, parks, churches, town halls, schools and universities, public buildings.[43] In a second step, the mark is identified by a

[42]Below a list of the 70 metropolitan areas, with the (approximate) number of map tiles covering the metropolitan area in parentheses: Barrow-in-Furness (70 tiles), Bedford (25), Birkenhead (80), Birmingham (300), Blackburn (80), Bolton (60), Bradford (100), Bristol (180), Burnley (45), Burton upon Trent (35), Cardiff (115), Carlisle (20), Castleford (10), Chester (40), Coventry (45), Crewe (20), Croydon (20), Darlington (10), Derby (60), Dover (40), Gateshead (15), Gloucester (60), Grimsby (50), Halifax (40), Huddersfield (50), Ipswich (50), Keighley (15), Kidderminster (30), Kingston-upon-Hull (140), Leeds (140), Leicester (80), Lincoln (30), Liverpool (300), London (600), Luton (15), Macclesfield (30), Manchester (260), Middlesbrough (15), Newcastle-upon-Tyne (150), Newport (70), Northampton (60), Norwich (35), Nottingham (350), Oldham (180), Peterborough (50), Plymouth (80), Portsmouth (60), Preston (80), Reading (25), Rochdale (40), Rochester (15), Sheerness (15), Sheffield (250), Southampton (40), Stockport (10), Stockton-on-Tees (40), Stoke-on-Trent (75), Sunderland (40), Swansea (35), Swindon (30), Taunton (50), Tynemouth (80), Wallsend (50), Walsall (30), Warrington (30), Wigan (40), Wolverhampton (50), Worcester (60), York (70).

[43]We complement this approach based on historical maps by using the English Heritage GIS Data provided by the Ordnance Survey, and geolocating monuments and listed buildings.

recognition algorithm as well as the associated identifier. The recognition algorithm then associates geographic coordinates to this identifier. In a third step, we match geographic coordinates and the information stored in Excel separate files (type of landmarks, and name). This information is used to generate atmospheric pollution but also distance to industrial chimneys (Figure 7) and distance to amenities.

We also use these series of Ordnance Survey maps to create polygons of land use for the 70 cities and their outskirts, and we define city borders as the minimum polyline surrounding built-up areas. These borders are used to construct the control variable *Share area (city)*, and are used for sample selection in Appendix Table A9.

Census data in 1817 and 1881 In this project, we have obtained the authorization from the Cambridge Group for the History of Population and Social Structure to get access to the digitized micro-census of England & Wales in 1881 (about 26 Million individuals). The following variables are available: a parish code, an address, gender, age, marital status, birthplace, occupation, size of the household, number of relatives, inmates, offspring and servants, non-relatives and visitors. We use the parish and occupational classification developed by Shaw-Taylor and Wrigley (2014), which is already harmonized between 1817 and 1881.

We rely on a quasi-census of male occupations drawing upon 2 million observations. The Church of England kept records of baptisms, marriages, and burials in Parish registers, and the occupation of the father was reported for each baptism. Accordingly, the sample only features males who had a child in the covered period. More than 10,000 such Anglican baptism registers from 1813 to 1820 were digitized, and Shaw-Taylor and Wrigley (2014) associate a 1881 parish and an occupational code (following the PST structure, see Wrigley, 2010) to each observation.

We also use on the income tax levied between 1799 and 1816, and collect the 1815 property-tax assessment published at the parish level. In order to finance the Napoleonic Wars, 6 Schedules were developed as part of a generalized Income Tax. We use Schedule A assessment (tax on land income) as a proxy for land value (or rent per acre).

Bomb damage in 1940–1945 We use scans of the "Bomb Census survey records, HO 193(55-65), 1940–1945" provided by the National Archives. Only the following cities of our sample are covered: Barrow in Furness, Bedford, Birmingham, Bristol (Bath), Coventry, Derby, Dover, Gateshead, Grimsby, Ipswich, Jarrow (Wallsend), Lincoln, Liverpool (tracings only), London, Luton, Manchester (tracings only), Middlesbrough, Newcastle upon Tyne, Norwich, Nottingham, Oldham, Peterborough, Plymouth, Portsmouth (Gosport), Sheffield, Southampton, York. These Bomb Census map tiles consist Ordnance Survey maps—Six-inch to the mile, drawn between 1919 and 1939, on which the *Ministry of Home Security Bomb Census Organisation* recorded damage sustained during bombing raids. The positions of bombs are marked by a red dot (see Appendix Figure A15).

We digitize bomb damage as follows. First, we superimpose these scans with original Ordnance Survey maps in order to geolocate each map tile. Second, we transform red dots into featured points, we draw a buffer of 10 meters around each point and associate them to the intersected LSOAs. Third, we define the dummy *Bombs* (used in Figure A16) as being equal to 1 if at least one bomb impact intersects with the LSOA.

Steam engines in 1700–1800 and waterways We use the directory of steam engines between 1700 and 1800, as computed by Kanefsky and Robey (1980) and revised by Nuvolari et al. (2011). We focus on 728 individual engines built in our cities of interest, of which about 400 are bought by collieries or textile mills. We use the location, the information about the owner and possible comments, and we precisely geolocate 547 of these steam engines.[44]

We then map these locations, group them into small geographic clusters ("historical industrial districts"), define the centroid of these clusters and generate pollution dispersion from these centroids as if they were industrial chimneys with uniform pollutant emissions. The resulting measure is the variable *Pollution (steam engines)* of Appendix Table A10. The variable *Pollution (waterways)* of Table 3 is constructed as follows. We collect navigable waterways in 1827, select the intersection of these polylines with the 1890 city borders (as computed by land use in Ordnance Survey maps), and locate hypothetical chimneys every 150 meters along the within-city waterways. We then generate pollution dispersion from these hypothetical chimneys using the same dispersion process as in the baseline and uniform emissions.

Recent data (1971–2016) *Aggregate censuses at the LSOA level (1971–2011)*— We collect small area statistics from the 1971, 1981, 1991, 2001 and 2011 Censuses. We construct the shares of low- and high-skilled workers from the 1-digit Socioeconomic categories; the shares of first-generation migrants from the "Country of birth"

[44]The following cities of our sample are covered: Birmingham, Blackburn, Bolton, Bradford, Bristol, Burnley, Carlisle, Castleford, Darlington, Derby, Gateshead, Hull, Keighley, Kidderminster, Leeds, Leicester, Liverpool, London, Macclesfield, Manchester (Salford), Newcastle upon Tyne, Northampton, Norwich, Nottingham, Oldham, Plymouth, Portsmouth, Preston, Reading, Rochdale, Sheffield, Stockport, Stoke on Trent, Sunderland, Tynemouth, Walsall, Warrington, Wigan, Wolverhampton, York.

variables; the shares of social housing and ownership using "Tenure and amenities" questions. We use the 2001 Lower Layer Super Output Areas (about 2,000 people per LSOA on average) as the main unit of analysis throughout the paper, and we sometimes use the Middle Layer Super Output Areas (about 10,000 people per MSOA on average) to clean for more granular fixed effects (Appendix Table A9) or to compute measures of neighborhood similarity (Figure A16).

House prices (1995–2015)—We rely on two datasets. First, we collect the HM Land Registry Transaction Data since 1995. This gives us an exhaustive register of all residential transactions in England and Wales. However, the data do not include other individual-transaction controls than the type of housing (detached, semi-detached, terraced, flat) and its age (new/old). Second, we use data from Nationwide, one of the largest mortgage provider in England and Wales, between 2009 and 2013. The Nationwide dataset includes a wide range of controls for property characteristics (e.g., the construction date, the number of bedrooms, bathrooms, garages, the square meters or heating facilities) but only covers 15% of Land Registry transactions. We use the average LSOA transaction prices in Table 6, and we use housing type as dependent variable in Appendix Table A11.

Building age—We collect use measures collected by the Consumer Data Research Centre that includes information on residential dwelling ages, grouped into three age bands. We use these measures as dependent variables in Appendix Table A11.

Education—We gather school outcomes for all primary schools from the Ministry of Education and generate LSOA measures of school supply (private schools, school value-added, teacher-pupil ratio, teacher salary, spending per student), school composition (disadvantaged pupils: defined as being either eligible for Free Schools Meals in the last six years; or looked after continuously for 1 day or more), or outcomes (the average GPA at Key Stage 2—primary education, pupils aged 7 to 11) for the period 2012–2013.[45]

Crime—We collect records of all criminal incidents in 2011 and their coordinates as reported by the police, and classify them into 4 categories: Anti-social behaviors including nuisance, vandalism, street drinking, littering, or vagrancy; Burglary; Drug-related crimes; and Violent crimes.[46] The number of crimes per inhabitants in each category is used as dependent variable in Appendix Table A11.

Deprivation indices—The English Indices of Deprivation (2010) are provided

[45] In order to collapse school-level indicators at the LSOA level, we proceed as follows. We compute the distance between every LSOA centroid and all the neighboring schools. We then aggregate all measures weighting each school by the inverse of the distance to the LSOA centroid.

[46] See `http://data.police.uk`. Note that we treat these incidents irrespectively of the outcome (court decision), and compute the number of such incidents in 2011 per 100 inhabitants.

by the Social Disadvantage Research Centre at the Department of Social Policy and Social Work at the University of Oxford. The main index is a weighted average of several sub-indices (Income, Employment, Health Deprivation and Disability, Education, Skills and Training, Barriers to Housing and Services, Crime, Living Environment Deprivation), some of which using primary data that are already directly included in our analysis (e.g., Key Stage 2 scores for the Education sub-index or crime occurrence for the Crime sub-index).[47] Further information on the composition and construction of sub-indices can be found in: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2010

Amenities—We construct contemporary amenities using the Ordnance Survey data on "Points of Interest" which contain the location of all public, education and leisure services. We then construct eight indices capturing the number of 1. Parks and recreation areas, 2. Theaters and museums, 3. Churches, 4. Hospitals, 5. National and local authorities, 6. Courts and police stations, 7. Transport infrastructure (bus and train stations), 8. Botanical gardens and zoos, per 100 inhabitants.

Current pollution—We collect SO2 pollutant concentration for the United Kingdom in 2010, as published by the Department for Environment Food & Rural Affairs (DEFRA). We use the *Current pollution* measure in Appendix Table A8.

[47]Two sub-indices may be interesting for our purpose, because they capture outcomes directly influenced by contemporary pollution. The Health Deprivation and Disability sub-index combines a measure of premature death with a morbidity/disability ratio, the emergency admission to hospital and the proportion of adults suffering from mood and anxiety disorders. The Living Environment Deprivation index captures housing quality but also air quality.

D Geolocating individuals in census data

This section describes the census structure, the fuzzy matching procedure, the clustering algorithm and a validation procedure.

Census structure There is a strong but imperfect relationship between census neighbors and true geographic neighbors that we clarify below. As we observe the parish, all our analysis will be for individuals of the same observed parish.

Let the *census identifier* i denote a transformation of the book/folio/line numbers in a systematic order. For each entry i, we can define a step function $f : i \mapsto f(i) \in \mathbb{N}$, monotonous in i.[48] The function f defines clusters among census entries. Let $n : i \mapsto n(i)$ denote the unobserved neighborhood for an individual entry i. We assume a monotonicity property for n reflecting that enumerators were recording households in a sequential manner: If $i < j < k$ and $n(i) = n(k)$, then $n(i) = n(j) = n(k)$. If two entries are in the same block n, all entries appearing between these entries also belong to the same block.

The monotonicity property is not fully sufficient to match households. Indeed, it does not allow us to observe the relationship between the values taken by blocks $\{n_j\}_j$ and census clusters $\{f_j\}_j$, and this is due to the fact that breaks in blocks cannot be observed. For instance, within a single parish, a list of entries can be:

id i	folio f(i)	block n(i)	break
1.	f_1	n_1	
$\vdots$	$\vdots$	$\vdots$	
45.	f_1	n_1	
46.	f_1	n_2	B_1
$\vdots$	$\vdots$	$\vdots$	
78.	f_1	n_2	
79.	f_2	n_2	B_2

As can be seen in the previous example, there are two types of breaks in the data, one associated with a change in blocks B_1 that cannot be observed and one associated with a change in books B_2 which is observed. The true measure of a geographic cluster (i.e., a neighborhood) is n and census cluster f is an observed, imperfect proxy. In what follows, we will describe our strategy as if census clusters were a perfect representation of geographic clusters.

[48] For instance, we can group lines of the same folio/book by groups of 10, or group all entries of the same folio together.

Fuzzy matching of addresses We clean addresses by deleting blanks, normalizing terms used to indicate types of roads (e.g., road, street, avenue, bow, park, square, cottage, villas, etc.) and separating the road denomination from the attributed name. We reduce the probability of georeferencing a census address incorrectly by limiting the pool of potential matches (the contemporary geolocated addresses, and monuments and listed buildings from the English Heritage GIS Data) to those which are located in the registered parish of the census observations.

The fuzzy matching procedure generates perfect matches for 20% of the total sample, and we match 30% of the total sample with precision 0.90 (at least 90% of the original string can be found in the matched address). The covariation among census entries in unmatched addresses is small which indicates that most of the matching error comes from idiosyncratic sources. However, there remains some covariation, e.g., when streets are not found in contemporary directories or when a very large "census household", e.g., a jail, a boarding school or a guesthouse, has a poorly reported address.

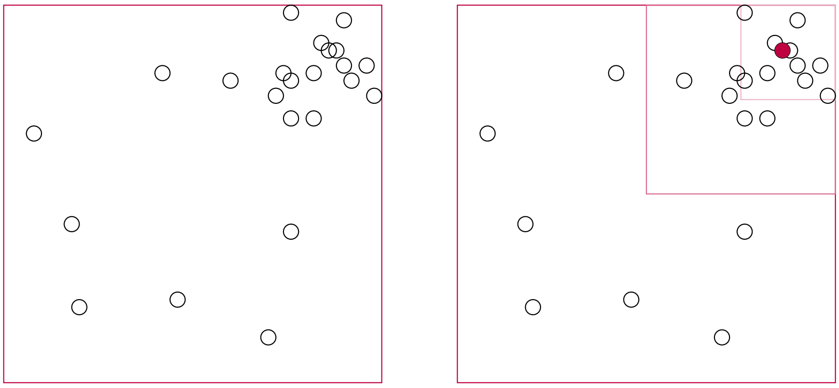
In what follows, we only keep matches with a score higher than 0.90 and consider the other cases as being unmatched. We describe in the following section how we account for potential errors in the already-matched household addresses and how we geolocate the remaining households.

Recognizing clusters and inference We infer the geolocation of all households of the same census cluster f from the geolocation of a subsample of households (with potential measurement error). We start with the sample of well-matched households and apply the following algorithm to detect geographic clusters.

1. Geolocate all georeferenced households.
2. Divide a parish into 4 equal regions, depending on their position relative to the maximum and minimum latitudes and longitudes in the sample.
3. Select the region with the largest number of observations, and temporarily drop the other observations.
4. Go back to point 2. with the newly selected region.
5. Stop after few iterations, calculate the average location of remaining households, and attribute this georeference to *all* households in the same census cluster f. We then overlay these newly-identified blocks with our consistent geographic units and attribute a unique LSOA identifier to each observation.

A graphical illustration of this algorithm is provided in Figure A11 with 2 iterations. Two (resp. three) iterations already divide a parish into 16 (resp. 64) small regions. The advantage of this process is twofold: it infers georeferences for unmatched households and it reduces noise in georeferences among already-matched units.

Figure A11. Finding geographic clusters among georeferenced households in the same census cluster.

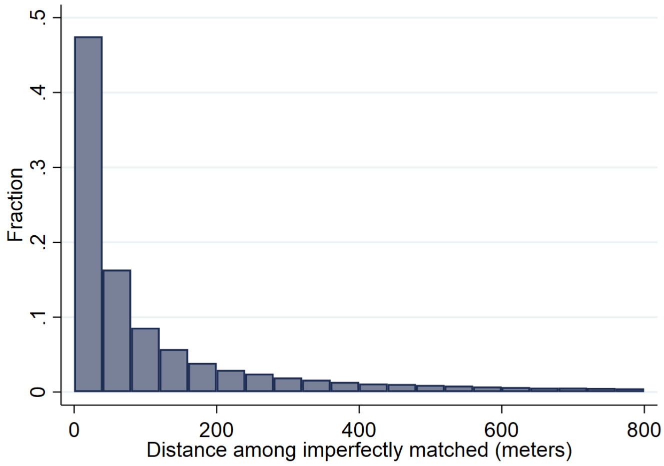


Measurement error and sensitivity analysis The previous methodology relies on two approximations. First, while census clusters reflect underlying geographic clusters, there remains a tension when aggregating entries into a census cluster. Having more households per census cluster raises the probability to detect a geographic location. There may, however, exist breaks within a book: the first household may be interviewed in another neighborhood than the last household, for instance if the latter are interviewed by another surveyor. In order to alleviate this concern, we repeat our algorithm by generating different census clusters (grouping 1, 2, 3, 4, 5 pages together, drawing new breaks) and we compare the resulting LSOA identifier under the different specifications. Second, the exact number of iterations in the previous algorithm or the 0.90 precision threshold to exclude poorly-matched addresses may matter. In particular, when two clusters coexist within the same group of households, the previous algorithm will select one of the two clusters and ignore the presence of the other. In order to identify these outliers, we keep track of the number of households located in the selected cluster and we generate a dummy equal to one when this number is lower than 0.50. These cases represent less than 5% of all observations.

We provide a formal validation of the clustering procedure in Appendix Figure A12: we draw a random validation sample of perfectly-matched households in

1881, accounting for 20% of all perfectly-matched households; we run the clustering algorithm using only the geo-location of perfectly-matched households outside of the validation sample; and we measure the distance between the newly-attributed coordinates and the actual coordinates (see Appendix Figure A12 for the distribution of such distance within the validation sample). About 90% of the validation sample is located within 50 meters of the previous location and about 95% is within 200 meters. An urban LSOA is typically between 0.5 and 1 square kilometer, which implies that almost all households of the validation sample are attributed to the same LSOA in both instances.

Figure A12. Validation of the clustering algorithm.



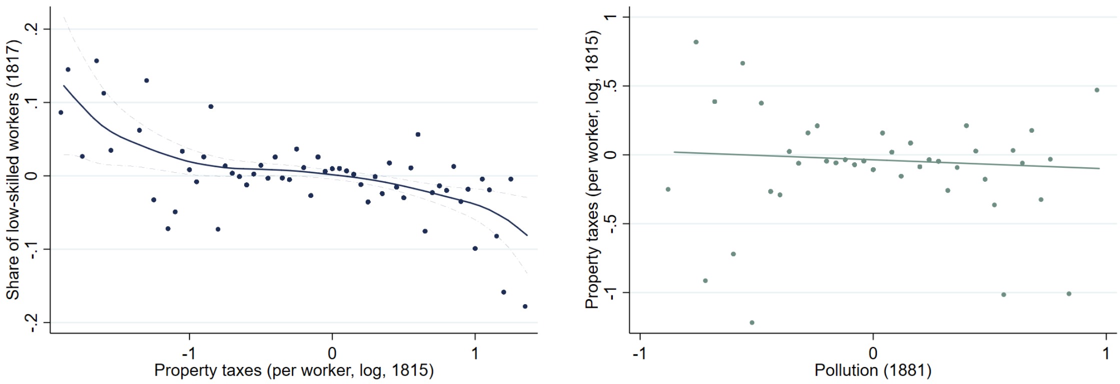
Notes: The Figure represents the distance between the baseline coordinates attributed to households and the coordinates attributed by the validation exercise for households within the validation sample and with non-zero distance (22% of the validation sample). About 78% of households in the test sample are attributed similar coordinates as in the baseline. About 90% of households are within 50 meters of their previous location, about 95% are within 200 meters.

E Sensitivity analysis

This subsection presents robustness checks around the baseline specification(s).

Balance tests and validation of the 1817 quasi-census The quasi-census of 1817 relies on the occupation of fathers as reported in baptism records of the Church of England over 1813–1820; a large share of the population may be under-represented, for instance older individuals or migrants from Catholic countries. We provide in the left panel of Appendix Figure A13 a cross-validation of the baptism records data: there is a strong correlation between the two proxies for neighborhood affluence around 1815. Numerous other validation tests are provided in Shaw-Taylor and Wrigley (2014).

Figure A13. Validation of the quasi-Census (1817) using property taxes at the parish level in 1815.



(a) Correlation between the share of low-skilled workers in 1817 and property taxes in 1815.

(b) Relationship between property taxes in 1815 and pollution.

Sources: The left panel represents the locally weighted regressions on all observations between the share of low-skilled workers in 1817 and the (log) property taxes per capita at the parish level (1815). The right panel represents the relationship between the (log) property taxes at the parish level (1815) 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 *lat./lon.* controls (see Table 2). We create bins of neighborhoods along past pollution and the dots represent the average (log) property taxes per capita within each bin. The line is a locally weighted regression on all observations. We restrict the sample to observations with residual pollution between −1 and 1 standard deviation(s).

Figures 8 and 9 provided suggestive evidence that within-city spatial inequalities before the energy transition to coal are unrelated to the later pollution exposure. In order to further reduce concerns about biasing effects from unobserved pre-existing neighborhood characteristics, we describe a more formal balance test in Appendix Table A4. Panel A mirrors Table 2 and shows the (absence of) correlation between atmospheric pollution and the 1817 share of low-skilled workers at the parish level.

We also exploit property tax data from 1815 to infer the average wealth at the parish level and run a similar balance test in Panel B of Appendix Table A4 (see the right panel of Appendix Figure A13 for visual evidence).

Table A4. Pollution and shares of low-skilled workers or wealth measures before pollution—balance tests in 1817.

Panel A: Share of low-skilled (1817)						
	(1)	(2)	(3)	(4)	(5)	(6)
Pollution	.0077	.0115	.0070	.0021	.0048	.0061
	(.0100)	(.0142)	(.0146)	(.0145)	(.0187)	(.0182)
	[.0715]	[.1063]	[.0654]	[.0201]	[.0451]	[.0567]
Observations	559	559	559	559	559	559
Fixed effects (city)	No	Yes	Yes	Yes	Yes	Yes
Controls (population)	No	No	Yes	Yes	Yes	Yes
Controls (topography)	No	No	No	Yes	Yes	Yes
Controls (amenities)	No	No	No	No	Yes	Yes
Controls (lat./lon.)	No	No	No	No	No	Yes
Panel B: Property tax (log, 1815)						
	(1)	(2)	(3)	(4)	(5)	(6)
Pollution	.3254	.4828	.2205	.1668	-.0283	-.0319
	(.0843)	(.1277)	(.0926)	(.0952)	(.1079)	(.1107)
	[.2623]	[.3892]	[.1777]	[.1344]	[-.0228]	[-.0257]
Observations	532	532	532	532	532	532
Fixed effects (city)	No	Yes	Yes	Yes	Yes	Yes
Controls (population)	No	No	Yes	Yes	Yes	Yes
Controls (topography)	No	No	No	Yes	Yes	Yes
Controls (amenities)	No	No	No	No	Yes	Yes
Controls (lat./lon.)	No	No	No	No	No	Yes

Robust standard errors are reported between parentheses. Standardized effects are reported between square brackets. Each cell is the result of a separate regression. The unit of observation is a parish. The set of *population* controls include the total population in 1817. The set of *topography* controls include the average, maximum and minimum elevations for the parish and the (inverse) distance to waterways as of 1827. The set of *amenities* controls include the (inverse) distance to the city hall, the (inverse) distance to parks, the parish share within the city borders in 1880, the parish area, the (inverse) distance to the closest heavy industry and the (inverse) distance to the closest light industry. *Lat./lon.* are the latitude and longitude of the parish centroid.

Chimney height and other pollution imprints We provide a sensitivity analysis of our working assumption on the effective height of the chimney. We only vary the chimney height as the exit velocity and the exit temperature affect the same crucial model input, i.e., the effective height of the smoke column in the atmosphere. We report in Appendix Table A5 the relationship between pollution, as emitted by chimneys that are 15 meters high and 40 meters high, and the share of low-skilled workers in 1881, 1971, 1981, 1991, 2001 and 2011. The estimates are similar to the baseline estimates of Table 2 and Table 4.

We use a simplified and transparent measure of pollution exposure in Appendix Table A6. We create a grid of equally-spaced points over each city, count the num-

Table A5. Pollution and shares of low-skilled workers—robustness to chimney height.

Panel A: Chimney height of 15m						
Share of low-skilled	1881	1971	1981	1991	2001	2011
Pollution	.0342	.0276	.0341	.0442	.0412	.0388
	(.0071)	(.0053)	(.0060)	(.0072)	(.0074)	(.0070)
	[.1352]	[.2169]	[.2431]	[.2359]	[.2535]	[.2222]
Observations	5,538	5,535	5,538	5,538	5,538	5,538
Panel B: Chimney height of 40m						
Share of low-skilled	1881	1971	1981	1991	2001	2011
Pollution	.0337	.0265	.0337	.0431	.0406	.0382
	(.0073)	(.0056)	(.0064)	(.0078)	(.0077)	(.0073)
	[.1329]	[.2081]	[.2404]	[.2302]	[.2495]	[.2188]
Observations	5,538	5,535	5,538	5,538	5,538	5,538

Standard errors are reported between parentheses and are clustered at the parish-level (as defined in 1881). Standardized effects are reported between square brackets. Each cell is the result of a separate regression. The unit of observation is a Lower Super Output Area. The set of extended controls include all controls of column 6 in Table 2.

ber of chimneys within a distance to each grid point and in four different quadrants (North-East, North-West, South-West, South-East), and collapse the measures at the LSOA level. The first column of Appendix Table A6 shows the correlation between the share of low-skilled workers in 1881 and the number of chimneys within 2 kilometers of the average household in the four different quadrants. Only chimneys located North-West and South-West matter—and equally so—, coinciding with prevailing downwind directions. In columns 2 and 3, we split the sample between Northern and Southern cities. Chimneys located South-West are relatively more predictive of deprivation in Northern cities, possibly reflecting the Southern winds prevalent in North England (see Figure 4).

We report a comprehensive sensitivity analysis relying on other pollution imprints in Appendix Table A7. The large number of chimneys across the city implies that a measure like the distance to the closest chimney—used as a control in the benchmark specification—may not fully capture the proximity to industrial centers. Instead, we construct an atmospheric pollution profile from existing chimneys under wind patterns that are symmetric in all directions (*symmetric pollution*). As shown in column 1 of Appendix Table A7, the placebo atmospheric pollution measure does not affect our estimates.[49] In column 2, we present a placebo pollution pattern that varies the emission intensity rather than wind patterns. Specifically, we assume low emissions for chimneys in polluting industries and high emissions for chimney

[49] As the measure of symmetric pollution is positively correlated with the share of low-skilled workers when we do not control for actual pollution, this finding indicates that (i) there are more low-skilled workers close to factories but (ii) they are mostly located downwind of factories (both features were already apparent in the descriptive Figure 7).

Table A6. Number of neighboring chimneys and shares of low-skilled workers in 1881.

Share of low-skilled workers	(1)	(2)	(3)
Number of chimneys (NE)	.0001	.0024	-.0072
	(.0080)	(.0102)	(.0122)
Number of chimneys (NW)	.0343	.0374	.0255
	(.0062)	(.0096)	(.0076)
Number of chimneys (SW)	.0332	.0422	.0215
	(.0079)	(.0099)	(.0115)
Number of chimneys (SE)	.0133	.0070	.0161
	(.0080)	(.0092)	(.0131)
Observations	5,538	2,741	2,797
Sample	All	North	South
Fixed effects (city)	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes

Standard errors are reported between parentheses and are clustered at the parish-level (as defined in 1881). The variable *Number of chimneys (XX)* is defined as the average number of chimneys—within 2 kilometers and in the *XX* quadrant—of a randomly drawn household within a LSOA. The *North* and *South* samples are defined relatively to the median latitude of city centroids (around the latitude of Wolverhampton). The unit of observation is a Lower Super Output Area. The set of extended controls include all controls of column 6 in Table 2.

in non-polluting industries.[50] Including this measure does not affect our estimates and we observe a negative and insignificant coefficient on the placebo pollution measure, which shows that using information on the related industry is critical. In column 3, we control for residential pollution, which has some predictive power: the standardized effect of residential pollution is about three times smaller than the effect of industrial pollution.

We repeat this exercise with the share of low-skilled workers in 2011 as the dependent variable in Appendix Table A8.[51] In this table, however, we also add a measure of contemporary pollution, i.e., the average yearly NO2 concentration measured by DEFRA in 2015. We find that contemporary pollution has a relatively small impact on neighborhood composition in 2011. Historical pollution is far more predictive of current spatial inequalities than current exposure to air pollutants.

Sensitivity to controls, fixed effects, sample selection and clustering In Appendix Table A9, we condition for city-specific spatial gradients with respect to production locations, we discuss the choice of fixed effects, clusters, and the robustness of the estimates to the exclusion of some regions. In Panel A, we control

[50] We rank industries by their coal use per worker (see Appendix Table 1), and attribute the smallest value to the most polluting industry, the second smallest value to the second most polluting industry etc.

[51] Appendix Figure A7 replicates, for 1991 and 2011, the rotation exercise performed in Figure 9, and shows that deprived areas are still situated within a narrow cone around prevailing winds originating from past industrial centers.

Table A7. Pollution and shares of low-skilled workers in 1881—other pollution imprints and residential pollution.

Share of low-skilled workers	(1)	(2)	(3)
Pollution	.0431	.0492	.0438
	(.0104)	(.0137)	(.0136)
	[.1701]	[.1940]	[.1729]
Symmetric pollution	-.0094	-.0071	-.0083
	(.0092)	(.0099)	(.0099)
	[-.0369]	[-.0281]	[-.0328]
Placebo industry		-.0110	-.0089
		(.0150)	(.0149)
		[-.0432]	[-.0351]
Residential pollution			.0141
			(.0047)
			[.0555]
Observations	5,538	5,538	5,538
Fixed effects (city)	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes

Standard errors are reported between parentheses and are clustered at the parish-level. The unit of observation is a Lower Super Output Area. The set of extended controls include all controls of column 6 in Table 2. *Symmetric pollution* is the counterfactual pollution exposure using a constant average wind along all wind directions, and the same pollution sources and emissions as in the baseline. *Placebo industry* is the counterfactual pollution exposure using the same dispersion as in the baseline, but changing emission intensity at origin. *Residential pollution* is the pollution generated by residential emissions. See Section 2 for additional details about the construction of these variables.

for latitude and longitude (column 1), distance to the town hall (column 2), and distance to heavy industries and distance to light industries (column 3), each time interacted with city-fixed effects. These sets of controls condition for within-city geography, allowing for different dependence in distance to the town hall or the main industries (e.g., capturing different commuting facilities across cities). The effect of one standard deviation in pollution remains stable between 3.6 and 4.5 percentage points in the share of low-skilled workers. In Panel B—column 1, we report the results of our baseline specification with parish-fixed effects (about 540 in our sample). We further expand our set of fixed effects in column 2 to electoral wards (1,440 in our sample) and in column 3 to Medium Lower Super Output Areas (1,850 in our sample). The estimates remain unchanged, even when identification comes from a within-MSOA comparison. In Panel C, we report standard errors clustered at three different levels, electoral ward, MSOA and city. In Panel D, we compute spatial standard errors that allow for spatial correlation along distance (following Conley, 1999)—instead of clustering standard errors at a certain administrative level (e.g., parish in the baseline). Standard errors increase by about 70% between the least and most conservative specifications, and by about 40% between our baseline analysis clustered at the parish-level and the most conservative choice.

Table A8. Pollution and shares of low-skilled workers in 2011—other pollution imprints, residential pollution and contemporary pollution.

Share of low-skilled workers	(1)	(2)	(3)	(4)
Pollution	.0364	.0380	.0413	.0405
	(.0078)	(.0108)	(.0113)	(.0113)
	[.2085]	[.2176]	[.2365]	[.2315]
Symmetric pollution	-.0010	-.0004	.0003	.0004
	(.0070)	(.0073)	(.0076)	(.0076)
	[-.0057]	[-.0024]	[.0018]	[.0021]
Placebo industry		-.0028	-.0041	-.0032
		(.0114)	(.0117)	(.0118)
		[-.0163]	[-.0236]	[-.0182]
Residential pollution			-.0087	-.0090
			(.0046)	(.0046)
			[-.0498]	[-.0517]
Current pollution				.0106
				(.0037)
				[.0605]
Observations	5,538	5,538	5,538	5,538
Fixed effects (city)	Yes	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes	Yes

Standard errors are reported between parentheses and are clustered at the parish-level. The unit of observation is a Lower Super Output Area. The set of extended controls include all controls of column 6 in Table 2. *Symmetric pollution* is the synthetic pollution exposure using a constant average wind along all wind directions, and the same pollution sources and emissions as in the baseline. *Placebo industry* is the pollution exposure using the same dispersion as in the baseline, but changing emission intensity at origin. *Residential pollution* is the pollution generated by residential emissions. *Current pollution* is the average yearly NO2 concentration measured by DEFRA in 2015. See Section 2 for additional details about the construction of these variables.

In Panel E, we estimate the baseline specification on alternative samples. We exclude Greater London in column 1, the North-West including Manchester and Liverpool in column 2, and the North-East in column 3. The estimates fluctuate around the baseline, but they remain large in all cases. In Panel F, we analyze the sensitivity of our results to the exclusion of suburbs and rural LSOAs. In columns 1 and 2, we exclude LSOAs outside of a range of 10 and 5 kilometers around the town hall, and we exclude LSOAs whose share of area within the 1880 city borders is equal to 0.5 in column 3. Even in the last instance with only 30% of our original observations, the estimate remains precisely estimated and slightly larger than in the baseline.

Alternative instrument The location decision of polluting industries in the early nineteenth century may be associated with future planned development of downwind neighborhoods. In this section, we suggest an alternative way to obtain exogenous variation in industry location, based on the historical settlements of industries. We isolate variation induced by the location of industrial districts *before* coal became a major energy source which would affect disproportionately downwind neighborhoods. To predict early industrial districts, we locate 543 early steam engines installed between 1700–1800, using data from Kanefsky and Robey (1980) and Nuvolari et al. (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. Appendix Table A10 report four sets of estimates, restricting the sample to urban neighborhoods within a distance of 10,000 meters or 5,000 meters of a textile mill, and with or without extended controls in addition to city fixed-effects.[52] The IV estimates are remarkably similar to the ones using waterways as an exogenous factor for industry location (see Table 3).

[52] During this time, steam engines were predominantly used in textile production and collieries. Our data show that collieries are most frequently located in the hinterland of few cities, e.g., Newcastle upon Tyne, Gateshead, Leeds or Sheffield. Since we are interested in industrial districts *within* cities, we restrict our sample to urban neighborhoods around historical textile mills.

Table A9. Pollution and shares of low-skilled workers in 1881—sensitivity to flexible geographic controls within cities, fixed effects, clustering and sample selection.

	Share of low-skilled workers (1881)		
	(1)	(2)	(3)
Panel A: Flexible controls			
Pollution	.0420	.0453	.0355
	(.0078)	(.0078)	(.0079)
	[.1657]	[.1789]	[.1401]
Observations	5,538	5,538	5,538
Controls (× City-FE)	Coordinates	Distance to hall	Distance to chimney
Panel B: Fixed effects			
Pollution	.0387	.0377	.0393
	(.0091)	(.0115)	(.0118)
	[.1527]	[.1490]	[.1551]
Observations	5,538	5,538	5,538
Fixed effects	Parish	Ward	MSOA
Panel C: Clusters			
Pollution	.0338	.0338	.0338
	(.0058)	(.0066)	(.0086)
	[.1332]	[.1332]	[.1332]
Observations	5,538	5,538	5,538
Clusters	MSOA	Ward	City
Panel D: Spatial standard errors			
Pollution	.0338	.0338	.0338
	(.0066)	(.0094)	(.0101)
	[.1332]	[.1332]	[.1332]
Observations	5,538	5,538	5,538
Cut-off (kms)	1	5	10
Panel E: Sample selection (excluding regions)			
Pollution	.0297	.0642	.0343
	(.0070)	(.0126)	(.0069)
	[.1172]	[.2536]	[.1353]
Observations	3,661	4,395	5,247
Excluding...	London	NW	NE
Panel F: Sample selection (excluding rural neighborhoods)			
Pollution	.0339	.0350	.0425
	(.0068)	(.0075)	(.0093)
	[.1340]	[.1383]	[.1677]
Observations	5,122	3,521	1,583
Sample	Hall<10km	Hall<5km	LSOA ∩ city>0.5

Standard errors are reported between parentheses and are clustered at the parish-level. Standardized effects are reported between square brackets. Each cell is the result of a separate regression. The unit of observation is a Lower Super Output Area. All specifications include all controls of column 6 in Table 2. In Panel A, we control for latitude and longitude interacted with city-fixed effects (column 1), distance to the town hall interacted with city-fixed effects (column 2), and distance to heavy industries and distance to light industries interacted with city-fixed effects (column 3). A *Ward* is an electoral ward (election for local councils), and there are 1,300 wards in our sample. A *MSOA* (Middle Lower Super Output Area) is the second smallest unit in the census, and there are 1,800 MSOAs in our sample. *London* is Greater London and includes 33 districts in addition to the City of London. *NW* is the north-western region while *NE* is the north-eastern region. In Panel F, we exclude (1) LSOAs outside of a range of 10 and (2) LSOAs outside of a range of 5 kilometers around the townhall, (3) LSOAs whose share of area within the 1890 city borders is lower than 0.5.

Table A10. Pollution and shares of low-skilled workers in 1881 (IV specification)—sensitivity analysis with another instrumental variable.

Panel A: First stage	Pollution			
	(1)	(2)	(3)	(4)
Pollution (steam engines)	.2760	.1700	.3331	.1937
	(.0493)	(.0325)	(.0599)	(.0400)

Panel B: Second stage	Share of low-skilled workers (1881)			
	(1)	(2)	(3)	(4)
Pollution	.0884	.0901	.0752	.0801
	(.0252)	(.0384)	(.0244)	(.0402)
	[.3488]	[.3558]	[.2970]	[.3160]
Observations	2,519	2,667	1,860	1,860
F-statistic	31.30	27.33	30.89	23.49
OLS coefficient	.0297	.0221	.0272	.0193
Sample	Mill<10km	Mill<10km	Mill<5km	Mill<5km
Fixed effects (city)	Yes	Yes	Yes	Yes
Extended controls	No	Yes	No	Yes

Standard errors are reported between parentheses and are clustered at the parish-level. Standardized effects are reported between square brackets. The top panel reports the first stage, and Kleibergen-Paap F-statistics are reported in the bottom panel. The unit of observation is a Lower Super Output Area. All specifications include all controls of column 6 in Table 2. The variable *Pollution (steam engines)* is the predicted pollution instrument which uses the location of steam engines between 1700–1800 as pollution sources. Steam engines proxy the center of historical industrial districts which were typically producing textiles. In columns 1 and 2 (resp. 3 and 4), we exclude LSOAs outside of a range of 10 (resp. 5) kilometers around a textile factory.

F Differences across neighborhoods in school supply, crime, housing quality and public amenities

The quantitative analysis developed in Section 3 is silent about the exact nature of the neighborhood effects that may operate, e.g., peer effects, inertia in the housing stock or the accumulation of durable public amenities constructed during the Industrial Revolution such as parks or public services. This section describes empirical facts about formerly polluted neighborhoods in 2011. While this analysis is not causal and cannot be used as hard evidence in favor of one particular channel of persistence, it helps understand the within-city distribution of consumption amenities and its relationship with past atmospheric pollution.

In order to understand the correlation between the within-city distribution of consumption amenities and past atmospheric pollution, we report the estimates for specification (S1) where we replace our benchmark indicator of neighborhood composition by measures of local amenities (see Appendix Table A11): (i) deprivation sub-indices (Panel A), (ii) a selected set of schooling and crime indicators (Panel B), (iii) characteristics of the housing stock (Panel C), and (iv) selected city amenities (Panel D). Formerly polluted neighborhoods are consistently ranked as more deprived areas across all sub-indices of deprivation. Note that the measures *Income*, *Employment* and *Education* are the most correlated with past pollution. These measures capture the incidence of low earnings, involuntarily exclusion from the labor market and a lack of attainment and skills in the local population. By contrast, the correlation between *Housing*, measuring the limited physical and financial access to housing and local services, and past pollution is quantitatively small.

We then exploit more precise measures of schooling quality and crime incidence in Panel B. While the presence of private schools and the school value-added are negatively correlated with past pollution, the effects are quantitatively small. More generally, we verify in unreported tests that measures of school supply (e.g., teacher-pupil ratio, teacher salary, spending per student) are not strongly correlated with past pollution. Instead, measures capturing directly or indirectly school composition (disadvantaged students or scores) are markedly different in formerly-polluted neighborhoods. Along the same lines, burglary, drug-related and violent crimes, that tend to happen in poorest areas, are more frequent in these formerly-polluted neighborhoods in contrast to anti-social behaviors (see columns 5 to 8).

Panel C reports the correlation between past pollution and house age (columns 1 to 4). Formerly-polluted neighborhoods are not more likely to have houses constructed before 1970, 1940 or 1900, as confirmed by the average year of construction for transactions recorded by Nationwide. However, the housing supply remains different in these areas: one standard deviation in past pollution is associated with a 5 p.p. higher prevalence of flats and a 2 p.p. lower prevalence of villas.[53]

Finally, as shown in Panel D, formerly polluted neighborhoods have more parks, recreational areas and transport facilities, and less hospitals, botanical gardens or conference centers but all estimates are small in magnitude. The demand for high-quality amenities in good neighborhoods may be counteracted by high land prices. The presence of parks and recreation areas in formerly polluted neighborhoods may not only be due to low land prices but also to former industrial sites being destroyed and reclaimed in the second half of the twentieth century.

[53]Note that controlling for rigid consumption amenities (e.g., provision of public services, Victorian housing stock or old private schools constructed during industrialization) does not affect the gradient between neighborhood composition and past pollution. The housing stock or the city structure does correlate with historical pollution but has limited predicting power on neighborhood composition.

Table A11. Past pollution, and deprivation measures, education and crime indicators, housing quality and amenities in 2011.

Panel A: Deprivation indices	Index	Income	Empl.	Educ.	Health	Housing	Crime	Environ.
Pollution	.0627	.0681	.0491	.0781	.0341	.0003	.0273	.0559
	(.0119)	(.0135)	(.0116)	(.0120)	(.0083)	(.0050)	(.0076)	(.0131)
	[.2435]	[.2407]	[.1783]	[.2704]	[.1346]	[.0012]	[.1133]	[.2271]
Observations	5,538	5,538	5,538	5,538	5,538	5,538	5,538	5,538
Panel B: Education and crime	Private School (KS2)	Student Scores (KS2)	Disadvantaged Students (KS2)	School VA (KS2)	Anti-social Behaviors	Burglary	Drug-rel. Crimes	Violent Crimes
Pollution	-.0014	-.0049	.0073	-.0000	.0042	.0314	.0092	.0654
	(.0027)	(.0012)	(.0013)	(.0001)	(.0044)	(.0084)	(.0021)	(.0120)
	[-.0110]	[-.1040]	[.0887]	[-.0007]	[.0327]	[.0925]	[.1186]	[.1854]
Observations	5,538	5,538	5,538	5,538	5,538	5,538	5,538	5,538
Panel C: Housing quality	Building 1900	Building 1970	Building 2000	Year of construction	Square meters	Bedrooms	Flats	Detached
Pollution	.0011	-.0227	.0008	1.293	-2.171	-.0337	.0714	-.0253
	(.0109)	(.0117)	(.0053)	(1.537)	(.7340)	(.0165)	(.0117)	(.0054)
	[.0043]	[-.1027]	[.0047]	[.0396]	[-.0866]	[-.0542]	[.2649]	[-.1761]
Observations	5,538	5,538	5,538	5,226	5,226	5,226	5,538	5,538
Panel D: Amenities	Parks	Entert.	Church	Hospital	Public	Justice	Transport	Botanical
Pollution	.0139	-.0144	.0155	-.0010	.0153	.0028	-.0001	-.0064
	(.0117)	(.0103)	(.0080)	(.0018)	(.0108)	(.0035)	(.0077)	(.0033)
	[.0317]	[-.0357]	[.0549]	[-.0125]	[.0389]	[.0242]	[-.0003]	[-.0392]
Observations	5,538	5,538	5,538	5,538	5,538	5,538	5,538	5,538

Standard errors are reported between parentheses and are clustered at the parish-level. Standardized effects are reported between square brackets. Each cell is the result of a separate regression. The unit of observation is a Lower Super Output Area. The set of extended controls include all controls of column 6 in Table 2. The deprivation measures are the ranks of an LSOA (0: least deprived, 1: most deprived) along the different composite sub-indices constructed with Census data, housing data, vacancies posted, schooling outcomes, the presence of public services etc. (see Section 1). KS2 stands for Key Stage 2 (age 7–11). Building 1900, 1940 and 1970 stand for the shares of dwellings constructed before 1900, between 1900 and 1970, and after 2000. Years of construction, square meters and number of bedrooms are constructed from the non-representative set of Nationwide transactions, while the shares of flats and detached houses are constructed using the exhaustive Land Registry of transactions. Panel D uses as dependent variables the number of (i) Parks and recreation areas, (ii) Theaters and museums, (iii) Churches, (iv) Hospitals, (v) National and local authorities, (vi) Courts and police stations, (vii) Transport infrastructure, (viii) Botanical gardens and zoos, per 100 inhabitants.

G Role of social housing (1971–2011)

One factor that has fostered residential segregation between 1971 and 2011 is the liberalization of social housing. In 1979, Thatcher offered social housing tenants the 'Right to Buy' their property, which endogenized the distribution of social housing from 1979 onward.[54] The liberalization removed support for low-skill occupations in otherwise desirable neighborhoods. As a result, some low-skilled workers chose to sell the now-valuable housing to high-skilled workers. We use the Census in 1971, 1981, 1991, 2001 and 2011 and extract a LSOA-specific share of households living in council housing, in owned properties, and the share of migrant households.

Table A12. Pollution and social housing/migrant shares (1971–2011).

Effect of pollution on ...	1971	1981	1991	2001	2011
Social housing	.0035	.0191	.0316	.0260	.0283
	(.0138)	(.0123)	(.0097)	(.0078)	(.0073)
	[.0136]	[.0653]	[.1278]	[.1190]	[.1404]
	.287	*.358*	*.297*	*.260*	*.232*
Owners	-.0021	-.0086	-.0251	-.0247	-.0312
	(.0077)	(.0089)	(.0074)	(.0065)	(.0065)
	[-.0082]	[-.0318]	[-.1035]	[-.1073]	[-.1399]
	.429	*.494*	*.580*	*.583*	*.535*
Migrants (New Commonwealth)	.0067	.0147	.0143	.0172	.0253
	(.0034)	(.0046)	(.0046)	(.0056)	(.0075)
	[.1030]	[.1730]	[.1718]	[.1812]	[.2101]
	.041	*.060*	*.064*	*.085*	*.128*
Migrants (Other)	.0004	-.0003	-.0008	-.0006	.0033
	(.0014)	(.0012)	(.0008)	(.0011)	(.0014)
	[.0081]	[-.0054]	[-.0116]	[-.0098]	[.0439]
	.034	*.035*	*.043*	*.053*	*.075*
Observations	5,534	5,538	5,538	5,538	5,538
Fixed effects (city)	Yes	Yes	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes	Yes	Yes

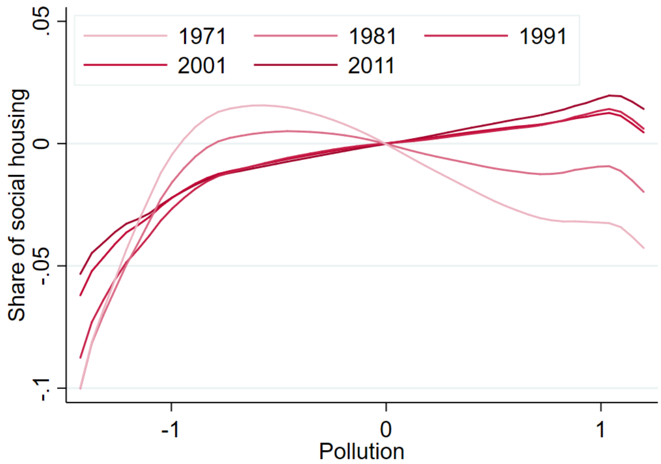
Standard errors are reported between parentheses and are clustered at the parish-level. Standardized effects are reported between square brackets. The average value for the explained variable is reported in italic. Each coefficient is the estimate for pollution in a separate regression. The unit of observation is a Lower Super Output Area. The set of extended controls include all controls of column 6 in Table 2.

Appendix Table A12 and Appendix Figure A14 show that, while social housing was weakly correlated with past pollution in 1971, it became increasingly present

[54]The United Kingdom initiated a program of social housing with the Housing of the Working Classes Acts (circumscribed to London in 1890 and extended to all councils in 1900). Council housing was the main supply of housing services for the working class, and it was typically managed by local councils. About 30% of urban households were living in council houses in 1971.

in formerly-polluted areas after 1979. More precisely, social housing appears to be distributed relatively uniformly across neighborhoods in 1971, but already aligns with historical pollution in 1991, reaching a steady-state afterwards. In parallel, the home-ownership rate decreases in areas that were formerly affected by coal pollution. We also report the correlation between past pollution and the share of immigrants in Appendix Table A12. The share of immigrants steadily increases in formerly-polluted areas with a sharp acceleration between the last two waves, which coincides with the rise in migration from poor countries. The conversion of social housing and the selective location of immigrants may be manifestations of residential segregation, but they may also have contributed to the persistence of neighborhood sorting.

Figure A14. Social housing (y-axis) and pollution (x-axis) across neighborhoods in 1971, 1981, 1991, 2001 and 2011.



Notes: The Figure represents the locally weighted regressions on all observations between the shares of social housing and our (standardized) measure of past pollution. We consider the residuals of all measures once cleaned by city Fixed-Effects, and the set of extended controls.

H Local factors driving the dynamics of persistence

A possible shortcoming of the quantitative analysis is that we cannot identify the different mechanisms underlying the estimated neighborhood effects. While a proper analysis of the different channels would go beyond the scope of the present investigation, we now provide indirect evidence on the possible impact of urban renewal policies.

"The Blitz" and local bomb damage We collect scans of Bomb Census maps covering about 30 cities of our sample, geolocate bomb damage (indicated by a red dot—see Appendix Figure A15), and define a dummy *Bombs* equal to 1 if at least one bomb impact has been recorded within the LSOA between 1940 and 1945. We then analyze bomb damages as shock to neighborhoods that induced changes in the local occupational structure.

Figure A15. Bomb Census maps—an example in Birmingham.

Sources: The National Archives, Bomb Census survey records—HO 193(55-65), 1940–1945. Red dots indicate the location of bomb damages.

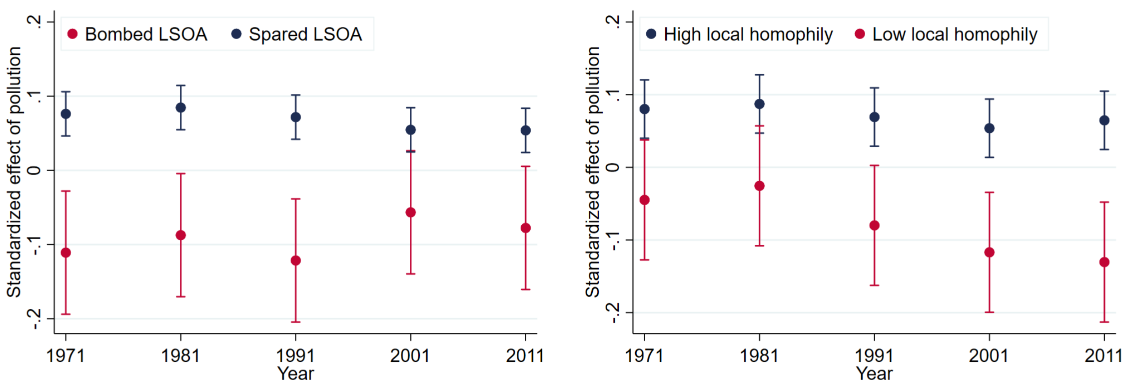
For this purpose, we run a (triple) difference-in-difference specification and identify the dynamics of persistence cleaning for location fixed effects. Letting i denote a LSOA and t a census wave ($t = 1881, \ldots, 2011$), we estimate the following equation:

$$Y_{it} = \sum_{\tau=1971}^{2011} \beta_\tau P_i \times \mathbb{1}_{\tau=t} + \sum_{\tau=1971}^{2011} \beta_\tau^b P_i \times \mathbb{1}_{\tau=t} \times Bombs_i + \sum_{\tau=1971}^{2011} \gamma_\tau \mathbf{X}_i \times \mathbb{1}_{\tau=t} + \nu_i + \delta_t + \varepsilon_{it} \tag{AS1}$$

where Y_{it} is the normalized measure of occupational structure and $\mathbf{X}_i$ is the same set of controls as in column 6 of Table 1. In this specification, β_τ captures the relative long-term impact of pollution in period τ (with respect to the impact in 1881, which is the omitted category), while β_τ^b captures the difference between bombed and spared LSOAs in this relative long-term impact of pollution.

We report the estimates of β_τ^b in Panel A of Appendix Table A13. As apparent in the first column, the relative long-term impact of pollution is, in 1971, $-.19$ lower for bombed LSOAs rather than for spared locations. This gap oscillates between $-.19$ and $-.13$ from 1971 to 2011. Appendix Figure A16 illustrates the differential long-term effects of pollution, β_τ^b and $\beta_\tau + \beta_\tau^b$, and their evolution over time. While the impact of pollution is higher in 1971 than in 1881 for spared areas, it is markedly lower in bombed LSOAs. These findings indicate that bomb damage had a smoothing effect on the local occupational structure, and removed the long-term gradient implied by historical pollution.

Figure A16. Persistence of neighborhood sorting—role of bombings and local homophily.



(a) Bombings. (b) Local homophily.

Notes: The left panel represents the difference-in-difference coefficients for historical pollution in bombed LSOAs, $\beta_\tau + \beta_\tau^b$, and spared LSOAs, β_τ. A difference-in-difference coefficient of 0.10 implies a relative standardized effect of pollution 0.10 higher than in 1881. Bombed LSOAs are LSOAs where at least one bomb impact has been recorded between 1940 and 1945. The right panel represents the difference-in-difference coefficients for historical pollution in LSOAs with high and low homophily. A high homophily LSOA is defined as a LSOA in which normalized pollution and the average normalized pollution of its neighbors within the same MSOA lie on the same side of the city average.

We consider these findings as possibly informative about the impact of urban

Table A13. Persistence of neighborhood sorting—role of bombings and local homophily.

Panel A: role of bombing	Share of low-skilled	Share of social housing
Pollution × Bombs × 1971	-.1870	
	(.0432)	
Pollution × Bombs × 1981	-.1719	-.0273
	(.0432)	(.0255)
Pollution × Bombs × 1991	-.1932	-.0619
	(.0432)	(.0255)
Pollution × Bombs × 2001	-.1112	-.0742
	(.0432)	(.0255)
Pollution × Bombs × 2011	-.1315	-.0800
	(.0432)	(.0255)
Observations	22,273	18,611
Fixed effects (LSOA)	Yes	Yes
Extended controls	Yes	Yes
Panel B: role of homophily	Share of low-skilled	Share of social housing
Pollution × Homophily × 1971	.1250	
	(.0396)	
Pollution × Homophily × 1981	.1126	-.0197
	(.0396)	(.0237)
Pollution × Homophily × 1991	.1491	-.0311
	(.0396)	(.0237)
Pollution × Homophily × 2001	.1707	-.0070
	(.0396)	(.0237)
Pollution × Homophily × 2011	.1950	.0033
	(.0396)	(.0237)
Observations	33,601	27,307
Fixed effects (LSOA)	Yes	Yes
Extended controls	Yes	Yes

Robust standard errors are reported between parentheses. All dependent variables are standardized. The unit of observation is a Lower Super Output Area×Census wave. The set of extended controls include all controls of Table 2—column 6 interacted with wave dummies. In Panel A, *Bombs* is a dummy equal to 1 if at least one bomb impact has been recorded within the LSOA between 1940 and 1945. In Panel B, *Homophily* is a dummy equal to 1 if the LSOA normalized pollution and the average normalized pollution of its neighbors within the same MSOA lie on the same side of the city average.

renewal policies and slum clearance.[55] Indeed, bomb damage may be considered exogenous at the local level once controlled for distance to industries. Following

[55] Slum clearances in England were numerous in the early twentieth century, following the Housing Act of 1930 and the Housing Act of 1936. However, it is not possible to extract exogenous variation as they were explicitly targeting deprived but well-located neighborhoods.

destruction, the government sometimes designated these bombed locations as redevelopment areas with newly-built social housing, which attracted poorer workers and may have had direct long-term spillovers on neighborhood composition (Redding and Sturm, 2016).[56] Instead, our main findings indicate that bomb damage induced redevelopment that worked against the persistence of segregation.

Local neighborhood tipping We now provide some empirical support for the theory of neighborhood tipping (Schelling, 1971). The idea is to capture how poorer LSOAs (or cells) evolve depending on the type of their immediate neighbors (or adjacent cells)—an exercise that resembles recent analyses of gentrification (Guerrieri et al., 2013; Baum-Snow and Hartley, 2020; Couture and Handbury, 2019).

We define a measure of homophily, i.e., individuals' taste for living among their own group type, as follows. *Homophily* is a dummy equal to 1 if the LSOA normalized pollution and the average normalized pollution of its neighbors within the same MSOA lie on the same side of the city's average pollution. High-homophily (resp. low-homophily) cells are thus surrounded by cells that are similar (resp. different) in pollution exposure.

We then run specification AS1 with the dummy *Homophily* instead of the dummy *Bombs*, and we report the results in Panel B of Appendix Table A13 and Appendix Figure A16. The relative long-term impact of pollution is, in 1971, .12 higher in high-homophily LSOAs, and this differential steadily rises to .19 in 2011. Historically-polluted cells that are surrounded by polluted cells (about 80% of the sample) are more likely to be locked in the same equilibrium. Figure A16 shows that capillarity across neighborhoods is particularly important after the Clean Air Act of 1968: while the correlation between occupational structure and pollution is already lower in low-homophily LSOAs around 1971, the gap doubles between 1971 and 2011.

The concentration of social structure has an impact on its dynamics. Air quality generated a spatially-correlated pattern in social structure within each city. Most highly polluted neighborhoods were adjacent to other polluted neighborhoods, thereby preventing occupational porosity across cells. This spatial correlation may be one component which drives the puzzling persistence of pollution effects in English modern cities. These findings may be informative about the reach of local spillovers. Urban planning and urban renewal policies may gain from targeting large areas all at once, as in Barcelona or London before their respective Olympic Games.

[56] Note, however, that we run specification AS1 for the normalized share of social housing in Panel A of Appendix Table A13, and we find that social housing is less correlated with historical pollution in bombed neighborhoods.