East Side Story: Historical Pollution and Persistent Neighborhood Sorting*

Stephan Heblich Alex Trew Yanos Zylberberg

February 18, 2018

Abstract

Why are the East sides of formerly industrial cities the more deprived? Using new individual-level census data together with new historical pollution patterns derived from the locations of 5,000 industrial chimneys and an atmospheric model, we show that this results from the persistence of neighborhood sorting that first emerged during the Industrial Revolution when prevailing winds blew pollution Eastwards. Pollution explains up to 15% of within-city deprivation in 1881. We characterize those areas where these equilibria persist to this day even in the absence of the initial pollution sources. A quantitative model identifies the role of non-linearities and tipping-like dynamics in such persistence.

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

*Heblich: University of Bristol, CESifo, IfW Kiel, IZA, SERC; stephan.heblich@bristol.ac.uk; Trew: University of St. Andrews; awt2@st-andrews.ac.uk; Zylberberg: University of Bristol, CESifo; yanos.zylberberg@bristol.ac.uk. We are grateful to 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, 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, Bristol, Cambridge, CERGE-EI, CUHK, Edinburgh, HSE St. Petersburg, Munich, the Ifo Institute, Odense, 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 to have Eastern suburbs that are notably poorer than Western suburbs. This observation is echoed in media stories about the East Side in London, New York or Paris and in popular culture (such as in the long-running BBC soap opera, *EastEnders*). This paper is the first to characterize empirically the existence and evolution of such a pattern, as well as its deep historical origins and its persistence to the present day. In particular, we show that this East-West gradient is a remnant of the atmospheric pollution which affected cities during the Industrial Revolution. We first focus on the nineteenth century and document the relationship between the distribution of air quality within English cities and neighborhood sorting. We find that pollution from historical factories account for 15% of the variation in neighborhood composition in 1881. In a second step, we turn to the post-pollution period and analyze the dynamics of neighborhood segregation. The effects of the now absent pollution are still felt to the modern day and there are non-linearities in the dynamics of persistence. In cities with high dispersion in environmental (dis)amenities, tipping forces anchored formerly polluted neighborhoods.

Although the impact of pollution on welfare in cities is often highlighted in policy debates, the long-run responses of economic agents are less well known. Providing such evidence is challenging. While measures of pollution on a fine spatial scale are readily available for today, until this paper we have lacked such data for an historical period. We present a novel strategy to model historical pollution within English cities at the end of the 19th century. This allows us to analyze the effect of pollution on neighborhood dynamics during industrialization as well as its effects in the long-run, even after the success of the 1968 Clean Air Act. As coal pollution created a legacy in the form of neighborhood sorting, a better understanding of these neighborhood dynamics holds important implications for the design of environmental policies in newly-industrialized economies. The analysis of the dynamics of persistence can also be instructive for policy makers who make decisions about urban policies aimed at reducing spatial inequalities in post-industrial economies. To this end, we develop a model of neighborhood tipping which provides evidence for non-linear forces in the dynamics of segregation, as in Card et al. (2008). Moreover, we quantify the role of local urban redevelopment and gentrification at the micro-scale (Schelling, 1971; Guerrieri et al., 2013).

In this article we present what is, to our knowledge, the first long-run analysis of the effects of pollution on the *internal* structure of cities through a process of residential sorting. Our empirical analysis combines a novel measure of pollution exposure *within* British cities during the time of the Industrial Revolution with unique panel data at the neighborhood level spanning nearly 200 years. Three methodological innovations help us generate these data. To model pollution exposure, we first develop a method to geolocate industrial chimneys from historical Ordnance Survey (OS) maps of the 70 largest metropolitan areas in England over the period 1880–1900. Second, we use the world-leading modelling system for atmospheric emissions (ADMS 5) that incorporates within-city information on terrain, wind directions, chimney dimensions, exit velocity and coal burning temperature to predict pollution dispersion from each individual chimney. Third, we develop a new method to assign individual entries in the 1881 census to low-level administrative units (for our purpose, the 2001 Lower Super Output Areas, LSOA). This method starts by geolocating a fraction of well-identified addresses; for the remaining cases, we develop a clustering algorithm which infers locations by exploiting the structure of census enumeration books. This allows us to study residential sorting within 70 metropolitan areas at the level of 5,500 LSOAs in total, instead of at the level of the 600 ancient parishes to which the data were originally mapped.

Our findings show a strong correlation between air pollution and the share of low-skilled workers in 1881. A pollution differential equivalent to the one between the 10% and 90% most polluted neighborhoods of Manchester would be associated with a gradient of 16 percentage points in the share of low-skilled workers.

The variation in pollution exposure results from a combination of the locations of pollution sources and a dispersion process involving wind patterns and topography. The ideal experiment to identify the causal impact of pollution on neighborhood sorting would be to randomly allocate a chimney within a city or a parish and then compare upwind and downwind neighborhoods with similar topography at the same distance from the chimney. In order to isolate such variation in the data, we proceed in three steps. First, we show that the correlation is robust to the addition of a large set of controls that capture the separate effects of the distribution of pollution sources (e.g., proximity to industrial chimneys, distance to amenities), topography (e.g., elevation), and wind patterns (e.g., latitude, longitude). To reduce concerns about (fixed) omitted variable biases, we also discard the existence of unobserved pre-existing amenities in polluted neighborhoods by looking at the correlation with neighborhood composition around 1815. There may be, however, dynamic omitted variation related to the proximity to industries. In a second step, conditioning our analysis on the distance to industrial chimneys, we analyze neighbourhood composition in all different directions relative to the chimney. This spatial differencing exercise shows excess deprivation north-east of each chimney. This pattern remains when we aggregate pollution across all chimneys: the share of low-skilled workers is markedly lower within a narrow corridor that coincides with the prevailing winds. There remains one important concern, that chimneys may have been selectively located upwind of poor areas. In a third step, we neutralize variation induced by the strategic distribution of pollution sources using instrumental variables. We instrument the pollution pattern induced by actual chimneys with two predicted pollution patterns that exploit exogenous locations of pollution sources. The first set of exogenous pollution sources draw from the insight that steam engines need water for cooling (Maw et al., 2012). Consequently, we exploit natural waterways in 1827 as exogenous location factor and locate pollution sources uniformly along these waterways to predict the actual pollution pattern. As we 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. Our second predicted pollution pattern exploits information on the location of steam engines installed between 1700 to 1800 (Kanefsky and Robey, 1980; Nuvolari et al., 2011). We consider the location of individual workshops with steam engines as proxy for the historical industrial districts within cities *before* industrial coal smoke might have influenced location decisions. The two-stage specifications with either instrument deliver similar qualitative results and slightly larger estimates.

Having established that pollution caused neighborhood sorting in the past, we focus on more recent years and analyze the dynamics of persistence between 1971 (just after the second Clean Air Act of 1968) and 2011.[1] At the beginning of this period, pollution from coal burning abruptly decreased. Nevertheless, we find a significant and relevant effect of historical pollution on the social composition of neighborhoods. The estimates in 1971, 1981, 1991, 2001 or 2011 are similar and all quantitatively comparable to those in 1881. Past pollution explains up to 20 percent of the observed neighborhood segregation whether captured by the shares of blue-collar workers and employees, house prices or official deprivation indices.

The dynamics of persistence between 1971 and 2011 further shows patterns of non-linearities, with some mean-reversion for intermediate values of within-city pollution and more inertia for neighborhoods with extreme values of within-city pollution. In order to quantify these non-linearities, we develop a stylized model of neighborhood sorting which extends Lee and Lin (2017). The location choices of high- and low-income individuals depend on consumptive amenities with some amenities being tied to the neighborhood composition. We estimate the model to match the

[1]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 abrupt drop in coal consumption.

dynamics over the period 1881–1971. The best fit exhibits tipping-like dynamics with large tail effects for the endogenous consumption amenity. The model predictions for the period 1971–2011 closely match the observed evolution of neighborhood composition and explain both the within-city differences in returns to the mean but also the observed differences between heavily and mildly polluted cities.[2]

Finally, we focus on the local factors driving the dynamics of persistence. We first exploit the "Blitz", i.e., the 1940–1941 German bombing offensive, and show that bomb damage and urban redevelopment sever the long-run local relationship between pollution and neighborhood sorting. Second, we provide evidence for local neighborhood tipping (Schelling, 1971) and quantify how the LSOA persistence in occupational structure may depend on neighboring units.

Our paper contributes to different strands of the literature. First, our work is related to Lee and Lin (2017) who look at exogenous natural amenities as driver of neighborhood sorting. Instead, we look at coal pollution as a temporary amenity. To the best of our knowledge, we are the first to show that the large, temporary pollution from industrial coal use modifies the spatial organization of cities in the long run.[3] Related papers that look at pollution-induced sorting today include Banzhaf and Walsh (2008); 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 (2016), 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. (2017) who document a similar correlation between pollution and residential choices *between* cities in contemporary China.

Second, we contribute to a burgeoning literature on the gentrification of historic centers in U.S. cities (Brueckner and Rosenthal, 2009; Guerrieri et al., 2013; Baum-

[2]Another interesting factor behind the persistence of neighborhood sorting is related to the liberalization of social housing and the 'Right to Buy' introduced by the Thatcher government in 1979. The model shows that this liberalization reinforced the persistence of spatial inequalities by lowering the existing barriers to neighborhood sorting.

[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 that were removed in the early 1960s (Brooks and Lutz, 2016).

[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., 2014), violent crime (Herrnstadt and Muehlegger, 2015; Heyes and Saberian, 2015), and health (Graff Zivin and Neidell, 2013; Anderson, 2015; Deryugina et al., 2016) which relates to our research. Closely related historical assessments of the effect of coal use on health include (Clay and Troesken, 2011; Barreca et al., 2014; Clay et al., 2016; Beach and Hanlon, 2016).

Snow and Hartley, 2016; Couture and Handbury, 2017) which builds on previous research on the dynamics of segregation and tipping points (Schelling, 1971; Card et al., 2008; Logan and Parman, 2016). After the sharp decline in industrial pollution in English cities, formerly polluted neighborhoods remain the poor parts of town. Our quantitative analysis points to non-linearities as the main driver of the dynamics of segregation: past a certain threshold, highly-polluted neighborhoods accumulate low amenities and attract low-income residents even after pollution has waned. We differ from most of the tipping-point literature in two dimensions. In our context, we mostly identify a social component behind segregation (in contrast to the literature on the United States mostly focusing on ethnic considerations). Moreover, we exploit a temporary disamenity to explain the initial spatial distribution of residents instead of permanent differences across neighborhoods.

Third, we make several methodological contributions to quantitative research in economic history. Our first contribution is to provide a methodology to digitize historical maps and fully exploit them as extremely valuable sources of information. 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. Our second contribution is to show the ability of state-of-the-art pollution models to estimate historical pollution. Our third and most important methodological contribution is to provide an algorithm that geolocates census entries in 1881 and 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 small share of well-matched neighbors.

The remainder of the paper is organized as follows. In Section 1, we develop a stylized model of neighborhood sorting. Section 2 briefly provides some elements of context. We detail our main data sources and methodology in Section 3, and our empirical strategy in Section 4. We analyze the relationship between neighborhood sorting and historical pollution in Section 5. Section 6 looks at the dynamics of persistence between 1971 and 2011 and relies on a quantitative, dynamic version of the model developed in Section 1. Finally, Section 7 briefly concludes.

1 Pollution and neighborhood sorting

In this section, we introduce a stylized framework to study the effect of pollution on neighborhood sorting. This static model is the foundation for a quantitative, dynamic version that we further develop in Section 6.

In the model, neighborhood sorting arises out of within-city differences in consumptive amenities.[5] Each neighborhood is made up of an interval of locations. The amenity level at a location is partly neighborhood-specific (such as air quality) and partly location-specific (such as the quality of the view). As in Lee and Lin (2017), willingness to pay rent in high amenity locations is relatively higher for high income types: in equilibrium, high income workers are housed in the best amenity locations, low income workers are in the remaining locations, and a difference in air quality between neighborhoods causes sorting of a portion of the high- (low-)skilled workers into the less (more) polluted neighborhood.

Environment A city is composed of two neighborhoods indexed $j \in \{W, E\}$ (West and East). The mass of land in each neighborhood is $\mu(\Omega(j)) = 1$. We assume that rent is collected by absentee landlords who lease land to the worker who will pay the most rent. The mass of workers is of measure 2. Workers are heterogeneous in their income, θ, and they are perfectly mobile. A fixed proportion γ are low-skilled workers with income θ^l; the remaining workers are high-skilled and have income $\theta^h > \theta^l$.

While the quantitative model in Section 6 will be dynamic, we assume here a static framework. Workers choose their location to maximize,

$$V(j, \ell) = A(j, \ell)c(j, \ell) \quad \text{subject to} \quad c(j, \ell) + R(j, \ell) = \theta, \tag{1}$$

where $A(j, \ell)$ is the amenity level in location ℓ of neighborhood j, $c(j, \ell)$ is consumption and $R(j, \ell)$ is rent. Since amenities increase the marginal utility of consumption, high-skilled workers will sort into the most attractive neighborhood locations.

The amenity at each location ℓ in each neighborhood j is made up of three components: a location amenity x, air quality a (at the neighborhood level), and an endogenous amenity d that will be inoperative in the present static framework,

$$A(j, \ell) = a(j) + x(\ell, j) + d(j). \tag{2}$$

Air quality and the endogenous amenity can differ across neighborhoods, but they are constant within neighborhoods. By contrast, the location factor varies within a neighborhood—different locations within a neighborhood have inherent differences in attractiveness (as in Davis and Dingel, 2014).[6] We assume that $x(\ell, j)$ is, in both

[5] By contrast, Redding and Sturm (2016) model the production side and estimate spillovers between neighborhoods.

[6] While within the same neighborhood all workers share the same air and can access the same endogenous amenities, some locations have an inherent advantage over others. One part of a neighborhood may have scenic views, for example.

neighborhoods, uniformly distributed over the interval $[0,1]$. In this static model, we normalize endogenous amenities to $d(j) = 0$ for $j = \{W, E\}$.

Equilibrium Since agents are perfectly mobile, workers of the same type obtain the same utility. Let utility to high-skilled workers be $\bar{V}^h$, and $\bar{V}^l$ to low-skilled workers. Without loss of generality, we normalize $\bar{V}^l = 0$; by consequence, the rent charged to a low-skilled worker is simply $R^l(j,\ell) = \theta^l$ for all (j,ℓ). Rent charged to a high-skilled worker, $R^h(j,\ell)$, is,

$$R^h(j,\ell) = \theta^h - \frac{\bar{V}^h}{A(j,\ell)}. \tag{3}$$

Landlords rent their land to the workers that pay the highest rent. Land is rented to a low-skilled worker if at (j,ℓ),

$$\theta^l \geq \theta^h - \frac{\bar{V}^h}{A(j,\ell)}. \tag{4}$$

Low-skilled workers sort into those locations with the worst amenities.

Equilibrium $\bar{V}^h$, and so $R^h(j,\ell)$, is obtained using equations (3), (4) and a land-worker clearing condition. In particular, $\bar{V}^h$ is such that the mass of land rented to low-skilled workers is equal to the total supply of low-skilled workers. Letting $I^l(j,\ell) = 1$ if location ℓ in neighborhood j is rented to a low-skilled worker, the land-worker clearing condition is,

$$\sum_j \int_{\ell \in \Omega(j)} I^l(j,\ell)\mathrm{d}\ell = 2\gamma. \tag{5}$$

Equations (4) and (5) imply that, in equilibrium, $\bar{V}^h$ is such that the 2γ locations with the lowest amenities across both neighborhoods host the low-skilled workers.

Proposition 1. *There exists a $\bar{V}^{h*} > 0$ such that the worker-land clearing condition is satisfied. High-skilled workers sort into those locations with amenities above $A^* = \bar{V}^{h*}/(\theta^h - \theta^l)$. Imperfect sorting at the neighborhood level can occur in equilibrium if amenity levels overlap.*

Proof. See Appendix B. □

Sorting and pollution Following Proposition 1, we denote $F(A)$ the cumulative density of land with amenity level less than or equal to A within the city, and we define $S^l(j)$ as the equilibrium share of low-skilled workers in neighborhood j.

In the absence of pollution, we have $a(W) = a(E) = 0$ and $d(W) = d(E) = 0$, so $F(A) = 2A$. The amenity level that satisfies (5) is where $A^* = \gamma$. The low-skilled share in neighborhood j is the share of land in the neighborhood with $A \leq A^*$, that is, $S^l(j,t) = A^* - \min_\ell\{A(j,\ell,t)\}$. Without pollution, neighborhoods are symmetric and $S^l(j,t) = \gamma$ for $j \in \{W, E\}$.

Pollution takes the form of emission of an air contaminant that causes air quality to decline. Pollution emitted in each neighborhood is ρ, but a Westerly wind blows a portion $\eta \in (0,1)$ of the pollution emitted in neighborhood W into the air of neighborhood E:

$$a(W) = -(1-\eta)\rho,$$
$$a(E) = -(1+\eta)\rho.$$

Lemma 1. *With imperfect sorting, pollution causes the East to have a larger proportion of low-skilled workers. More intense pollution causes more sorting.*

Proof. See Appendix B. □

The impact of pollution is depicted in Appendix Figure A1. The disamenity causes equilibrium rents paid by high-skilled workers to increase compared to the benchmark without pollution. Since the 2γ locations with the lowest amenities are now disproportionately in the East, the East has a larger share of low-skilled workers.

In our empirical exercise, we will provide evidence on the spatial relationship between pollution and the share of low-skilled workers at the peak of industrial pollution, relying—as in the model—on the asymmetric dispersion of pollution implied by wind patterns.

2 Historical context

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. Per capita growth rates did not accelerate until after 1830 (Crafts and Harley, 1992). The transition to coal as a dominant energy source occurred only after the 1840s.[7] This late energy transition is reflected in Appendix Figure A2: there is a sharp acceleration of coal consumption between 1850 and 1910 and a stabilization until 1960. The early twentieth century saw a consolidation of industry with employment peaking at 46% in 1950 (Crafts, 2014). Thereafter it declined, most rapidly in the 1980s when state-owned industries were privatized. The decline in coal consumption

[7] 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.

slightly preceded the massive de-industrialization. The Clean Air Acts of 1956 and 1968 introduced regulations that penalized, among other things, the emissions of grit, dust and 'dark smoke' and placed minimum height restrictions on chimneys. These Acts led industry to shift 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 Appendix Figure A2, these regulations had an immediate and marked impact on coal consumption.

The heavy reliance on coal between 1850 and 1950 generated unprecedented concentrations of sulphur dioxide, which scarred cities and their surroundings.[8] The negative impact of atmospheric pollution is captured in a well-known case of microevolutionary 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* between 1970 and 2000 (Cook, 2003).

In parallel to the structural transformation of the economy, the end of the eighteenth century also saw a rapid growth of population in cities and the migration of workers out of rural hinterlands into the emerging industrial cities (Shaw-Taylor and Wrigley, 2014). As shown in Williamson (1990) and the Appendix Figure A2, the growth of cities peaked in the 1830s and then 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.[9]

In our empirical exercise, we will observe: (i) urban composition in 1817, before the acceleration in coal consumption and around the end of the rural migration to urban centers; (ii) atmospheric pollution and urban composition around 1880–1900, slightly before the peak in coal consumption; and (iii) urban composition between 1971 and 2011, after the abrupt decrease in atmospheric pollution.

[8] Previous contributions had conjectured a relationship between historical pollution and neighborhood sorting: "In Manchester [...] prevailing and strongest winds [blow] from the South West. This meant that when the dense sulphurous smoke left Manchester's tall chimneys it usually moved North East, and this was to have a marked effect on the shaping of the city. [...] The poorest city dwellers were forced to live amongst the mills and factories in north-easterly districts [...] the better-paid among Manchester's working classes might at least escape the worst of the smoke." (Mosley, 2013)

[9] 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.

3 Data

This section describes the construction of the atmospheric pollution around 1880–1900 and neighborhood composition in 1817, 1881 and 1971–2011. We first explain how we identify industrial chimneys and generate the associated pollution imprint. We then describe our matching algorithm to geolocate the 1881 Census. Finally, we provide important descriptive statistics.

In this paper, we rely on various data sources including: the Ordnance Survey (OS) maps—25 inch to the mile (1842–1952); a quasi-census in 1817, property tax collection in 1815 and the 1881 micro-census; Bomb census maps (1940–1945); a directory of steam engines installed between 1700 to 1800 (Kanefsky and Robey, 1980; Nuvolari et al., 2011); census aggregate data (1971–2011); the English Indices of Deprivation (2010); the Nationwide and Land registry data for recent house transaction prices; and the National Pupil Database in 2012–2013. We provide an exhaustive description of these sources and constructed variables in Appendix C.[10]

3.1 Construction of the Air Pollution measure

Our strategy to generate a georeferenced air pollution map for 70 metropolitan areas covered by OS maps can be summarized as follows. In a first step, we iterate over 5,500 georeferenced map tiles and mark each chimney with a unique identifier. We use a recognition algorithm to locate the marks and extract each chimney's coordinates along with the associated identifier. In a second step, we predict atmospheric dispersion of polluting particles from each individual chimney, weighting by the industry-specific reliance on coal. In a third step, we consider a relevant geographic unit, the Lower Super Output Area in 2001, and overlay all chimney-specific pollution imprints to generate a unique air pollution measure for each geographic unit. We describe these three stages in more detail below.

Identifying chimneys We rely on OS maps to identify chimneys and factories. These maps come at a 25 inch:1 mile scale, by far the most detailed topographic mapping that covers all of England and Wales between 1880 and 1900. The maps contain details on roads, railway, rivers, canals, public amenities, the outline of each building and their use. Most useful for our purposes, these maps also outline, in a sign of the fastidiousness of Victorian mappers, a clearly marked location of factory chimneys. Symbols are either a small rectangle with an inner circle or a

[10]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.

large white circle, drawn to scale. In most maps, a *Chy* or *Chimney* is written to help identify these symbols. These variations in symbols and sizes prevent us from directly using a recognition algorithm (two examples of symbols are shown in Figure 1). Instead, we go through all map tiles and mark chimneys (about 5,000 in total) with a recognizable symbol X and a unique numeric identifier.

An example of the chimney-identification is provided in Appendix Figure A3. On this map fragment, four different chimneys can be identified.[11] The red symbol X, located in the center of a chimney, is identified by a recognition algorithm which, together with the projection provided by the Ordnance Survey, allows us to geolocate each chimney. An identifier, e.g., *00007*, follows the sign. The advantage of such a process is that information on industries can then be retrieved after the recognition algorithm has located a chimney and stored the associated identifier. For instance, the chimney *00007* belongs to *Eastbrook Dye Works*.

We restrict our analysis to the 70 largest cities in England at the beginning of the nineteenth century, as evident from the 1907 Census of Production. These cities constitute a quasi-exhaustive snapshot of industry and cover between 60% and 66% of the total population between 1801 and 2011.

Dispersion modelling Atmospheric dispersion is calculated using the *ADMS 5* dispersion modelling software.[12] This model is an augmented version of the basic Gaussian air pollutant dispersion equation known as the Gaussian-Plume model. *ADMS 5* models atmospheric dispersion under a large spectrum of meteorological conditions, provides reliable pollution estimates in coastal areas and incorporates the impact of temperature and humidity. Moreover, it accounts for complex terrain and the changes in surface roughness. Since industrial chimneys during the Industrial Revolution were at a much lower altitude than modern chimneys, pollution dispersion was heavily influenced by surrounding topography.

The *ADMS 5* model requires a large number of inputs. First, *ADMS 5* uses complex meteorological information for each city. We use contemporary 10-year statistical meteorological data as provided by the Met Office for the different cities in our sample, thereby neglecting small changes in prevailing winds related to climate fluctuations between the 19th century and today. Appendix Figure A4 illustrates the meteorological information for Northern England and Southern England. The "wind roses" show wind provenance and intensity. As expected, winds blow mostly from West/South-West. Moreover, we see that wind is less predictable in Northern

[11] On this map tile, chimneys are indicated with a plain white circle and the word *Chimney*.

[12] See http://www.cerc.co.uk/environmental-software/ADMS-model.html.

England generating on average more disperse air pollution measures.[13] Second, the model requires complex terrain data and convective meteorological conditions on land. We use the current terrain height and roughness, which affect wind speed and turbulence for cities with high gradients.[14] Finally, *ADMS 5* requires information on the emission source. Atmospheric dispersion modelling 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, most being lower than 25 meters. Moreover, the exit velocity and temperature were also lower than today. To incorporate this, we set chimney height to 25 meters and assume an exit velocity of 4 m/s and an exit temperature of 120 degrees Celsius.

We also model pollution related to domestic emissions. To this end, we assume domestic chimneys to be uniformly distributed within the city borders (as recognizable in OS maps) and we use the same meteorological and topographic inputs. However, domestic chimneys are lower, so pollution patterns are more localized.

Aggregation and some descriptive statistics Atmospheric dispersion models are additive. Total Air Pollution is calculated as the sum of pollution concentrations computed from each separate chimney. To account for sectoral differences in coal use, we employ the map information on the industrial site associated with each chimney and define the following industrial categories: Brick factories, Foundries, Chemical factories, Mining, Breweries, Tanneries, Food processing, Textile production, Paper production, Shipbuilding, Wood processing, and Other manufactures. We match these categories with national information on industry-specific coal use per worker (Hanlon, 2016) and weight the chimney-specific pollution imprint using this industry-specific coefficient. More precisely, we construct the estimated pollution emission E_i from a chimney in industry i 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 (L_i/Ch_i is thus the average number of workers per chimney in industry i). We report the estimated

[13] Under stable conditions and high chimneys, the wind carries pollution far from the origin source while pollution is most intense at the origin under unstable conditions (i.e., wind does not have a prevailing direction). Note that our benchmark measure uses an average of these different conditions over the past 10 years.

[14] In Appendix Figure A5, we show the differences in pollutant dispersion implied by topography in a city with high gradients, i.e., Oldham. Topography and land cover play little role in flat terrains.

pollution emission per industry, E_i, in Appendix Table A1.

We finally collapse our data at the level of 2001 Lower Super Output Areas to assign a pollution measure to administrative units. Figure 2 displays the industrial sources of pollution for Manchester (left panel) and the resulting aggregate Air Pollution (right panel). We can see that the pollution cloud tilts toward the East. Appendix Table A2 provides another illustration of the within-city variation in air pollution that can be used as an overidentification test. It reports a sample of deposits collected in Manchester by the First Annual Report of the Sanitary Committee on the Work of the Air Pollution Advisory Board, 1915.[15] We observe a very large variation across neighborhoods for both measures, illustrating that distance to chimneys, topography and wind directions generate very large within-city dispersion in pollution. Reassuringly, the estimated pollution strongly correlates with the deposit measure (correlation of 0.92).

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 LSOAs. Appendix Figure A6 shows that about 10% of our sample 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 our sample LSOAs—mostly in Manchester, Oldham and Liverpool—have indices of pollution above the peaks recorded in contemporary Beijing ($40\mu g/m^3$).

3.2 Geolocating individuals in census data

In order to measure neighborhood composition at a disaggregated geographic level, we use individual records from the 1881 census which hold information on the structure of households as well as, importantly, the address, age, sex, and occupation of its members. In this section, we briefly outline our methodology for allocating households interviewed in the 1881 census to contemporary administrative units. A detailed description can be found in Appendix D.

The intuition behind our methodology is the following. There are two indicators of household location: a geolocated parish variable and an unreferenced address. While the parish variable is consistently referenced, the address is inconsistently reported (surveyors use abbreviations and misspelling is frequent) and poorly digitized (e.g., due to handwriting). However, there exists another source of information in the 1881 census that has, to the best of our knowledge, not been exploited so far: individual surveyors were given blocks to visit on census day and they filled in enu-

[15]This first report happened to be the last one as well, such that these numbers are the only available elements of comparison for our pollution estimates.

merator books while visiting these neighborhoods. As a result, there is a strong clustering among census entries.[16] So if we locate a fraction of households, we can infer the georeferences of unmatched entries given (i) their location in the census books and (ii) their well-matched neighbors. In this way, we can assign individual records to smaller spatial units *within* parish boundaries.

Address matching In the transcription of the 1881 census enumerators' books, we observe the book number, folio number, and page number in addition to the already-exploited census variables.[17]

To implement our clustering analysis, we need to geolocate a non-negligible fraction of households in our sample. For this purpose, we carefully clean historical addresses by deleting blanks, normalizing the terms used for the types of roads (e.g., road, street, avenue, bow, park, square, cottage, villas, etc.) and we create a similar pool of contemporary geolocated addresses and listed buildings. We then run a fuzzy matching procedure between the pool of census addresses and the pool of geolocated addresses within the same parish of registration. We achieve a perfect match for 20% of the total sample, and we match 30% of the total sample with sufficient precision (90% of the original string is found in the matched address).[18]

Clustering algorithm A precise description of the algorithm is provided in Appendix D and we only discuss its main steps here. In a first step, we define a *cluster id* based on the book, page, and folio numbers for each record. This *id* will relate a census entry to its census neighbors. In a second step, we focus on the sample of well-matched households within each *cluster id*, analyze the cloud of located addresses, and identify the major cluster of points, its centroid, and the associated geographic unit. In a third step, we attribute this geographic unit to all entries with the same *cluster id*, including entries that were not matched during the fuzzy matching procedure. We repeat this algorithm with different cluster definitions, compare the resulting LSOA identifier under the different specifications, and select the most frequent LSOA identifier.

[16]Along the same lines, Logan and Parman (2016) exploit the structure of the 1880 U.S. census enumeration to create segregation measures based on the race of "census neighbors".

[17]These variables are: parish, address, surname, first name, relationship to head of household, marital status, gender, age, occupation, place of birth and disabilities.

[18]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) finally changes in street names, e.g., red-light districts. The first two sources of error are the most common.

Occupational measures We use "The Occupational Structure of England and Wales, c.1817–1881" (Shaw-Taylor and Wrigley, 2014) which cleans baptism records over 1813–20 to reconstruct a quasi-census of male occupations around 1817. Individuals can be linked to 834 parishes as defined in the 1881 Micro-census. For recent waves (1971–2011), we use area-weights to map census enumeration districts into 2001 Lower Super Output Area (LSOA) to generate persistent geographic units between census waves. The census data provide consistent measures of occupation, housing, education level and country of origin for all these years.

One drawback is that, at this level of disaggregation, we do not directly observe income, arguably the best proxy for the social composition of neighborhoods within cities. Instead, we observe 3-digit occupational information in the recent censuses and rely on a similar classification (PST system of classifying occupations; see Wrigley, 2010) for the two 19th century censuses (1817 and 1881). There are many ways to proxy for income based on occupational structure. For instance, one could predict the LSOA's income from average national wages by occupation and rentiers' income. However, such inference would require strong assumptions especially regarding the relative wage per occupation across cities. In order to make our analysis more transparent, we rely on a proxy based on the raw data, i.e., the LSOA's share of low-skilled workers among the working population.[19]

For 1817 and 1881, we first collapse 500 occupational categories into 10 categories. We then define low-skilled workers as Unemployed, Disabled, Unskilled and Semi-Skilled workers. Managers, Gentlemen, Rentiers, Clerks, and Manual Skilled workers are classified as high-skilled workers. Finally, we assign Farmers to a separate category and we drop Soldiers from our analysis. In order to refine our measure, we restrict our sample to individuals with the lowest possible measurement error, i.e., males between 25 and 55.[20] This decomposition brings about 60% of low-skills, 30% of high-skills and 10% of farmers in 1881 (78% of low-skills, 12% of high-skills and 10% of farmers in 1817) in our 70 metropolitan areas.

For 1971–2011, the occupational categories are already classified into 1-digit categories: Managers; Professionals; Associate Professionals; Administration; Manual Skilled; Care; Sales; Processing; and, Elementary. We replicate our classification in the main categories with two modifications. We group the first 3 categories as high-skills and the remaining 6 as low-skills to harmonize shares of low-skills be-

[19] We can alternatively normalize this share by the share of low skilled workers in the city. Doing so does not affect our results. Further note that we will use low-skilled occupations and low-skilled workers interchangeably in a slight abuse of notation.

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

tween 1881 and 1971–2011. Clerks and Manual Skilled workers are thus classified as low-skills, which brings about 62% of low-skills, 38% of high-skills in 2011. We drop the category "farmers" as it is almost non-existent among our urban LSOAs.

3.3 Descriptive statistics

We start by exploring in Figure 3 the assumed relationship between the share of low-skilled workers in 1881 and pollution sources in the raw data. Units of observation are the 675,000 block $\times$ chimney couples where a block is (i) a Census cluster of households with the same geolocation in 1881 and (ii) located within 2 kilometers of the chimney (there are 100,000 such blocks). The left panel of Figure 3 displays the average share of low-skilled workers in 1881 (adjusted by the number of observations including the same households such that all households are given the same weight), 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 the relatively high costs of commuting. Strikingly, conditional on distance to the pollution source and amenities in 1881 (distance to canals, town hall, theatres, hospitals, parks, churches, schools, universities, guild hall, mills, and elevation), 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 3, 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 more restrictive regression specifications in the next section.

Table 1 provides descriptive statistics at the level of our baseline units of observation. Within a buffer of 20 kilometers around the centroids of our 70 cities, the clustering process associates about 5 million active male workers in 1881 to $5,538$ LSOAs. As these LSOAs are 2001 census units, we can associate contemporary measures to all $5,538$ observations which will constitute our baseline sample. 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 (neighborhood composition in 1817, 1881 and 2011) and the baseline controls accounting for topography (elevation and distance to waterways), amenities (distance to the town hall, parks, heavy industries, light industries, share of LSOA within the city borders and area), and direction (latitude and longitude). Some of these controls 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. Interestingly, they are on average 2,000 meters further to the West than highly polluted neighborhoods.

In the last three columns of Table 1, we provide, for each variable, a decomposition of the variance within and between cities. A very large share of the variance in pollution is within cities. Our empirical strategy hinges on within-city variation and is mostly orthogonal to between-cities variation. It is thus reassuring to find such pollution dispersion across LSOAs of the same metropolitan area.

4 Empirical strategy

Benchmark specification To estimate the impact of pollution on neighborhood sorting within cities, we run a simple difference specification at the LSOA level in $t = 1881, \ldots, 2011$.

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

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

where Y_{it} is a measure of occupational structure. The simulated 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 simultaneously be a local (dis)amenity that affects 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 (employment in 1881, distance to waterways, to heavy-industry chimneys, to light-industry chimneys, to the town hall and to 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 property tax in 1815, which capture possible fixed neighborhood amenities. δ_c are city fixed effects.

We further exploit the interaction between the distribution of pollution sources and the dispersion process by considering counterfactual dispersion processes, e.g., the pollution dispersion generated by the same pollution sources but rotated wind patterns. One can think of this procedure as a decomposition of the interaction between location and dispersion to isolate 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 3).

Another 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 occupational structure in some neighborhoods. In robustness checks, we show a balance test for 1817, i.e., before the rise of industrial coal pollution, and provide identification at a more granular level by including fixed effects at the level of 1881 parishes or electoral wards.

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

IV specification In order to clean the variation in pollution from variations due to non-random industry location, we exploit exogenous geographic placements of industries to predict pollution imprints. In other words, we interact an exogenous variation underlying the choice of industry location with the atmospheric dispersion due to wind flows and topography and use the predicted pollution as our instrument.

We suggest two different ways to obtain exogenous variation in industry location. In a first specification, 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). To employ this insight, we locate hypothetical chimneys in intervals of 150 meters along natural waterways in 1827, before the rise of coal as the main energy source, and construct a predicted pollution instrument by constructing air pollutant emissions (without industry weights) from these predicted sources.[21] This natural geographic placement of chimneys is not susceptible of being selectively placed upwind of poor neighborhoods because of their emissions. However, the variation correlates with proximity to waterways which may itself affect the attractiveness of a neighborhood. We thus control separately for distance to the waterway and we exclude neighborhoods within a distance of 250 meters or 500 meters respectively.

In a second specification, we isolate variation induced by the historical location of industrial districts *before* coal became the major energy source that affected downwind neighborhoods. To predict early industrial districts, we identify the location of 543 steam engines within our sample of cities between 1700–1800 using data from Kanefsky and Robey (1980) and Nuvolari et al. (2011). During this time, steam

[21] The Appendix Figure A7 describes our approach. In panel (a), we see the cities of Manchester (left) and Oldham (right) with the associated 1827 natural waterways. Panel (b) displays the natural geographic placement of chimneys along canals and panel (c) the resulting air pollution. Finally, panel (d) shows the simulated historical pollution.

engines were predominantly used in textile production or collieries. Since we are interested in industrial districts *within* cities, we restrict our sample to urban neighborhoods within a distance of 10,000 meters (5,000 meters) of a textile mill.[22] Using early steam engine locations to proxy the centers of the historical industrial districts, we assume uniform air pollutant emissions and use the atmospheric dispersion model to create a predicted pollution instrument. As before, we separately control for distance to the nearest industrial chimneys and only rely on an upwind/downwind pattern in pollution around each industrial district.

We then use the following first-stage specification to instrument the simulated historical 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 one of the two predicted pollution instruments. 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 $\{\delta_c, d_c\}$ are city fixed effects.

5 Historical pollution and neighborhood sorting

In this section, we document a negative correlation between air pollution and neighborhood income as proxied by the share of low-skilled workers. The negative correlation is both economically and statistically significant at the peak of pollution in 1881: pollution explains 15% of the social composition across neighborhoods of the same city. While we control for important neighborhood characteristics in our benchmark specification (e.g., distance to main amenities and pollution sources, or neighborhood characteristics in 1817), we provide robustness checks on one potential shortcoming of our benchmark approach: The non-random location of industries. We first show a balance test, i.e., that atmospheric pollution is not correlated with the 1817 neighborhood average income. Second, we condition our analysis on neighborhoods being on the same ring around a factory, and isolate pollution variation from the direction of prevailing winds. Third, we run IV specifications to isolate exogenous variations in chimney locations. Finally, we present sensitivity analyses with more granular fixed effects, different sample selections, additional controls for amenities, and other outcomes in 1881.

[22] Our data show that collieries are most frequently located in the hinterland of few cities, e.g., Newcastle upon Tyne, Gateshead, Leeds or Sheffield.

Benchmark results In Table 2, we report the estimates for our baseline specification (S1) with $t = 1881$. 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 occupations between cities.[23] 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 our topography controls (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 0.033 and the 95%-confidence interval is [.020, .047]. One additional standard deviation in air pollution increases the prevalence of low-skilled workers by 3.3 percentage points, which is slightly less than 15% of a standard deviation in their prevalence across LSOAs. A differential in pollution equivalent to the one between the first and last deciles in Manchester would be associated with a differential of 16 percentage points in the share of low-skilled workers.

Figure 4 illustrates the estimated relationship between the shares of low-skilled workers 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 in 1881 on the same 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 flattens at both extremes, i.e., for very high and very low within-city pollution levels.

[23] The fixed effects also capture the negative correlation between coal-based pollution and city size, as documented in Hanlon (2016).

[24] As stated in Section 3, 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 the returns to agriculture).

[25] The first column of Appendix Table A3 reports the coefficients on all covariates.

Extensions and validity checks One threat to identification is that controls may not fully account for the potentially non-random and strategic location of industries. In this subsection, we address this issue and present other robustness checks.

First, we account for fixed LSOA characteristics in Table 3 (a "balance test"). Panel A of Table 3 mirrors Table 2 and shows the correlation between atmospheric pollution and the 1817 neighborhood average income as proxied by the share of low-skilled workers at the parish level. In all specifications, the coefficient is not different from zero statistically and economically. The (non-)relationship between pollution and the 1817 occupational structure is also shown in Figure 4. We also use property tax data from 1815 to infer the average wealth at the parish level and run a similar balance test in Panel B of Table 3. This approach reduces concerns about biasing effects from potentially unobserved pre-existing neighborhood characteristics. However, the location decision of polluting industries in the early nineteenth century may have been associated with future city development. We tackle this issue in the next set of robustness checks.

Second, we generate counterfactual pollution imprints from actual industry locations but alternative air pollution profiles to show that our estimates are not reflecting the mere distance to factories. We generate an index of pollution constructed from running the *ADMS 5* model on existing chimneys but with a *rotated* wind profile in steps of 30 degrees around each source, and we estimate its conditional correlation with the share of low-skilled workers. Figure 5 illustrates the results for the years 1817 and 1881 where we center the figure around the actual pollution pattern between 0 and 30 degrees. In 1817 (Panel a), *before* the rise of coal pollution, we observe virtually no correlation between the pollution measure at any degree and the share of low-skilled workers. In 1881 (Panel b), *after* pollution became a relevant disamenity, we observe a pronounced, bell-shaped pattern with the peak around the actual pollution pattern at 0 degrees. As we rotate prevailing winds, the estimated relationship loses significance and turns negative. The fuzzy relationship between 0 and 30 degrees may be due to measurement error.[26] In order 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.

We report additional sensitivity analysis with counterfactual pollution imprints

[26] 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 average for our meteorological conditions possibly ignoring differential pollution exposure and wind patterns across seasons or hours of a day.

in Table 4. We start from the baseline specification (column 6 of Table 2) and add pollution from a wind profile rotated 180 degrees (*Mirror* pollution). Next, we control for the density of chimneys around a neighborhood. The large number of chimneys across the city implies that a measure like the distance to the closest chimney—that we use as a control in the benchmark specification—may not fully capture the proximity to industrial centers. Instead, we construct an alternative atmospheric pollution profile from existing chimneys under a static wind profile, symmetric in all directions (*Static* pollution). One can think of this measure as a comprehensive distance decay measure. As shown in columns 1 and 2 of Table 4, these counterfactual atmospheric pollution measures do not affect our estimates.[27] In column 3, 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 in non-polluting industries.[28] Including this measure does not affect our estimates and we observe a negative and insignificant coefficient on the placebo pollution measure.[29] In column 4, we control for residential pollution as predicted by the location of private residential buildings, which has a small predictive power and again does not affect our main estimate.[30]

Third, we present in Table 5 the results of our IV strategy (S2), that uses 1827 natural waterways as a source of exogenous variation for chimney location (columns 1 and 2) and the location of steam engines as proxy for the historical center of industrial districts within the cities (columns 3 and 4). The first stage is strong in both cases. Both instruments generate 2SLS estimates which tend to be larger than the OLS estimates—also reported in Table 5. One additional standard deviation in air pollution increases the prevalence of low-skilled workers by about 7 percentage points in the specification with 1827 natural waterways, and 8 percentage points in the specification with steam engine locations. The results suggest that strategic location decisions do not seem to bias our results since we find evidence for downward-biased OLS coefficients. One explanation could be measurement error or possibly the difference between the local average treatment (relying on small

[27]As the measure of static pollution is positively correlated with the share of low-skilled workers when we do not control for actual pollution, these findings indicate 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 3).

[28]We rank industries by their coal use per worker (see Appendix Table A1), and attribute the smallest value to the most polluting industry, the second smallest value to the second most polluting industry etc.

[29]In an unreported robustness check, we verify that a measure of air pollution based on a uniform weight for each chimney as opposed to weights proportional to industries' coal consumption leads to similar results.

[30]This finding is to be expected since pollution from private homes is more homogeneously distributed across the city.

variations in possibly highly-polluted neighborhoods) and the average treatment. Following this exercise, we focus on OLS coefficients in the remainder noting that they might be a conservative lower bound.

Fourth, we provide sensitivity analyses for the following elements of our baseline specification: fixed effects, clustering, sample selection, controls and outcomes. In Appendix Table A4, we discuss the choice of fixed effects and clusters, and the robustness of the estimates to the exclusion of some regions. In Panel A—column 1, we report the results of our baseline specification with parish-fixed effects (about 540 in our sample) instead of city-fixed effects. We further expand our set of fixed effects in column 2 to current electoral wards (1440 in our sample) and in column 3 to Medium Lower Super Output Areas (1850 in our sample). The estimates remain unchanged even when identification comes from a within-MSOA comparison. In Panel B, 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. In Panel C, 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 Appendix Table A5, 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 townhall, and we exclude LSOAs whose share of area within the 1880 city borders is equal to 0 and lower than 0.5 in columns 3 and 4. Even in the last case (with only 30% of our original observations), the estimate remains precisely estimated and slightly larger than in the baseline. In the first column of Appendix Table A10–Panel A, we show that the relationship between atmospheric pollution and the 1881 neighborhood occupational structure is also unaffected by the inclusion of controls for the (potentially endogenous) local amenities, i.e., the number of parks, schools, theaters, museums, churches, hospitals per 100 inhabitants at the LSOA level.

Finally, we consider other outcomes in Appendix Table A6, with the 1881 share of all low-skilled workers including females and the 1881 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 all 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 (the Irish Potato Famine does not drive our findings).

6 Dynamics of persistence after the Clean Air Acts (1971–2011)

This section focuses on the relationship between historical atmospheric pollution and neighborhood composition in recent years and shows that the effect is of similar magnitude in 2011, almost 60 years after the first Clean Air Act and the subsequent drastic reduction in coal pollution. In order to quantify non-linearities in the dynamics of persistence, we develop and estimate a quantitative model à la Lee and Lin (2017). We then analyze the role of bombings during the 1940–1941 German bombing offensive—as a way to induce urban redevelopment—and we study the influence of homophily across neighborhing units in sustaining the occupational structure.

6.1 Historical pollution and contemporary neighborhood segregation

Persistence We expand on our previous analysis of neighborhood sorting to recent waves in 1971, 1981, 1991, 2001 and 2011. Table 6 reports the slopes between the shares of low-skilled workers and historical pollution. One additional standard deviation in historical air pollution increases the prevalence of low-skilled workers by 2.5 to 3.8 percentage points without a clear time pattern, and the standardized effects range between .19 and .23. A differential in pollution equivalent to the one between the first and last deciles in Manchester is still associated with a differential of 16 percentage points in the share of low-skilled workers, thereby explaining the social gradient between West and East often evoked in popular culture.[31]

In order to visualize non-linearities in the persistence of neighborhood sorting, Figure 6 displays the relation between shares of low-skilled workers in 1881 (dashed line), 1971, 1981, 1991, 2001 and 2011 (plain lines) and the historical pollution disamenity that stopped after the 1968 Clean Air Act. As apparent, we observe a reversion to the mean for intermediate values of within-city pollution. By contrast, segregation patterns remain constant at the extremes, i.e., around one standard deviation above or below average within-city pollution. This pattern would be consistent with the existence of tipping points leading to a higher persistence in neighborhoods with the most extreme pollution exposure.

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.[32] The liberalization removed support for low-skill occupations

[31]The persistence of the relationship between historical pollution and neighborhood composition cannot be mechanically attributed to the collapse of industries in the former *cottonopolis*, given that our estimates are identified within cities.

[32]The United Kingdom initiated a program of social housing with the Housing of the Working

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. Appendix Table A7 and Appendix Figure A8 show that, while social housing was weakly correlated with past pollution in 1971, it became increasingly present 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 A7. 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.

Robustness checks We now present further specifications that probe the robustness of the long-run relationship. First, Appendix Table A8 presents variations of the outcome measure. Panel A estimates the relationship between past pollution and a variety of deprivation indicators. One additional standard deviation in air pollution increases the income, employment, education and crime deprivation subindices by .10 to .30 standard deviations. These findings are confirmed by more direct measures of school quality and crime (see Panel B): student scores are lower in formerly-polluted neihborhoods, mostly driven by school composition, and violent crimes are much more frequent. As apparent in Panels C and D, housing quality varies along past pollution, but not building age nor the presence of amenities.[33] These dimensions of endogenous neighborhood amenities should be reflected in housing demand. In Appendix Table A9, we use transactions in England and Wales as recorded by Land Registry between 2000 and 2011 and Nationwide between 2009 and 2013 and we run a hedonic regression. The first columns of Appendix Table A9 report estimates for (log) house prices, one without house controls and one control-

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.

[33] In Appendix Table A10, we show that the relationship between past pollution and the current occupational structure is robust to controlling for the composition of local amenities, but the correlation decreases when controlling for indicators of housing quality, education or crime. Appendix E discusses these results in greater length.

ling for house characteristics. We find that one additional standard deviation in past pollution is associated with a price drop of about 10% to 11%. Controlling for house types reduces these estimates to 6-8%.[34] We provide an illustration of these large effects in Appendix Figure A9. Overall, these findings show the multi-faceted effects of historical pollution on neighborhoods' social composition, and its impact on neighborhood quality.

Second, we account for distance to former industry locations and potential reverse causality. Appendix Table A11 adds counterfactual pollution imprints with alternative air pollution profiles and pollution emissions, as well as contemporary pollution emissions that are not coal-related.[35] Appendix Figure A10 replicates, for 1991 and 2011, the rotation exercise performed in Figure 5, and shows that deprived areas are situated within a narrow cone around prevailing winds. Finally, Appendix Table A12 displays the 2SLS estimates for the occupational structure in recent years, and shows, as in Table 6, that there are no signs of reversion to the mean.

6.2 A quantitative and dynamic model of sorting

In Section 1, we laid the foundations of the quantitative model in a static framework of neighborhood sorting. We now extend the model to a dynamic framework where the persistence of sorting is rationalized by an endogenous amenity, which can be thought of as an index of "neighborhood quality".[36]

Workers are infinitely-lived and choose their location in each period to maximize,

$$V(j,\ell,t) = A(j,\ell,t)c(j,\ell,t) \quad \text{subject to} \quad c(j,\ell,t) + R(j,\ell,t) = \theta,$$

where $A(j,\ell,t) = a(j,t) + x(\ell,j) + d(j,t)$ and t is a calendar year. The location factor $x(\ell,j)$, constant over time, captures the fixed LSOA effect while $d(j,t)$ is an endogenous amenity that encompasses persistent neighborhood effects.

[34] An additional standard deviation in past pollution is also associated with a 5-8% decrease in the number of transactions (columns 5 and 6).

[35] Note that contemporary pollution has a relatively small impact on neighborhood composition in 2011 (5% of a standard deviation) and does not affect the predictive power of past pollution.

[36] Durlauf (2004) and Rosenthal and Ross (2015) are excellent overviews of the range of neighborhood effects affecting residential choices. If income levels differ, neighborhoods could accumulate amenity differences. These effects may include differences in school quality (Durlauf, 1996) or in the age of the housing stock (Rosenthal, 2008). Persistence could also work through peer effects. In this case, workers would simply have a preference to live among other workers in the same income group (Guerrieri et al., 2013) or ethnic group (Card et al., 2008). In the context of education, a peer effect may work via the presence of good role models (Benabou, 1993). Finally, peer and income effects could operate differently if a neighborhood composition crossed some threshold. Such tail effects would underpin the existence of poverty traps (Durlauf, 2004). Note that, in the model, workers can move freely across neighborhoods in any period, such that the persistence of sorting can only derive from the endogenous amenity.

The purpose of the quantitative model is not to posit micro-foundations for persistent neighborhood effects, but rather to estimate their structure from the data. We assume that $d(j,t)$ follows an AR(1) process with persistence $1-\delta$, and we allow for two types of endogenous perturbations. The first perturbation is a continuous neighborhood effect, $e(j,t)$, which increases in neighborhood j average income relative to the city-wide average income, $\bar{\theta}(j,t)$. Motivated by the possibility of tail effects, we model a component, $b(j,t)$, that reduces the attractiveness of neighborhoods beyond a threshold level of low-skill share. The endogenous amenity, $d(j,t)$, is thus defined for $t>1$ as,

$$d(j,t) = (1-\delta)d(j,t-1) + e(j,t) - b(j,t), \tag{6}$$

with a constant initial endogenous amenity across neighborhoods, $d(j,1)=d$.[37]

We consider the following parametric assumptions for the endogenous perturbations. As our framework has symmetric properties, amenities only matter through the implied difference between East and West neighborhoods and we will, without loss of generality, only load these neighborhood effects to the West neighborhood. The continuous perturbation is a continuous function of the neighborhood income (relatively to the city),[38]

$$e(j,t) = \phi_1^e \left[\bar{\theta}(j,t-1) - 1\right]^{\phi_2^e}, \tag{7}$$

The tail effect captures a discontinuity in neighborhood income, and detracts from the local amenity according to,[39]

$$b(j,t) = \phi_1^b \left[1 - \bar{\theta}(j,t-1)\right]^{\phi_2^b}, \tag{8}$$

The constants $\phi_1^e \geq 0$, $\phi_2^e \geq 0$, $\phi_1^b \geq 0$, $\phi_2^b \geq 0$, $\delta \in [0,1]$ and $\bar{S} > \tilde{\gamma}$ are unknown parameters to be estimated.

Before proceeding, Lemma 2 shows that, if the initial pollution caused one of the endogenous amenity perturbation to operate, then the sorting of neighborhoods persists. If neither channel operates, there is no sorting once pollution ceases.

Lemma 2. *Pollution can cause the accumulation of amenity differences and persistent sorting.*

[37] We do not find significant differences in amenities along past pollution in 1881.

[38] This can also be written in terms of the share of low-skill in j, $e(j,t) = \tilde{\phi}_1^e \left[\gamma - S^l(j,t-1)\right]^{\phi_2^e}$ where $\tilde{\phi}_1^e \equiv -\phi_1^e \left[\frac{(\theta^l - \theta^h)}{\gamma\theta^l + (1-\gamma)\theta^h}\right]^{\phi_2^e}$ and where the restriction $S^l(j,t-1) > \gamma$ applies.

[39] Again, in terms of S^l, $b(j,t) = \tilde{\phi}_1^b \left[S^l(j,t-1) - \gamma\right]^{\phi_2^b}$ where $\tilde{\phi}_1^b \equiv -\phi_1^b \left[\frac{(\theta^l - \theta^h)}{\gamma\theta^l + (1-\gamma)\theta^h}\right]^{\phi_2^b}$.

Proof. See Appendix B. □

We identify the model in the data using the within-city residuals of low-skill shares between $t_p = 1881$ and $t_c = 1971$ as well as the within-city residuals of atmospheric pollution for the 5,538 neighborhoods that we treat as independent observations. Let $p(j)$ be the normalized pollution in neighborhood j at time $t_1 = t_p$, i.e., $p(j) = \eta\rho$ in the East and $p(j) = -\eta\rho$ in the West. We can connect the model to the data by writing down the change in the share of low-skilled workers between $t_1 = t_p$ and $t_2 \geq t_c$ in a neighborhood j as a function of $p(j)$. This is the sum of the reversion that results from the pollution now absent at t_2 and the persistence in the accumulated $d(j,t)$,[40]

$$S^l(j,t_2) - S^l(j,t_1) = \underbrace{-\alpha p(j)}_{reversion} + \underbrace{\text{sign}\{p(j)\} \cdot d(j,t_2)/2}_{persistence} \tag{9}$$

where $\alpha > 0$ captures the empirical sensitivity of sorting to the normalized pollution.

With an initial pollution effect ($\alpha > 0$) but without any neighborhood effects ($\phi_1^e = \phi_1^b = 0$), the model predicts full convergence—Equation (9) is linearly decreasing in $p(j)$. In this case, the initial pollution causes sorting but there is later full reversion to the mean. If instead $\phi_1^e, \phi_2^e > 0$, the continuous neighborhood effect would act to solidify the initial sorting. If there was less than average historical pollution in a neighborhood, this effect would dampen the long-run reduction in the share of low-skilled workers. Moreover, if $\phi_1^b, \phi_2^b > 0$, then we may see a discontinuous effect around $\bar{S}$. Those neighborhoods most affected by pollution may see an additional long-run increase in the share of low-skilled workers.

Estimation We use the data to estimate the parameters of the endogenous amenities e and b as well as the persistence parameter δ. In Table 7, we report the model parameters that are first selected to match the data. We rely on Williamson (1980) for data on income inequality in nineteenth century England: We set the ratio of high income to low income at two.[41] Finally, we use the correlation between within-city residuals in low-skills and atmospheric pollution in 1881 to calibrate the sensitivity of low-skill share to pollution.

We estimate the remaining six model parameters (ϕ_1^e, ϕ_2^e, ϕ_1^b, ϕ_2^b, $\bar{S}$ and δ) using the method of simulated moments and we select those parameters that yield the best fit of the model to the observed change in low-skill share over the period 1881–1971.

[40] Lemma 1 shows that the share of low skill at the start of pollution is $S^l(j,t_p) = \gamma + p(j)$. Lemma 2 shows that the post-pollution share of pollution is $S^l(j,t_c) = \gamma + \text{sign}\{p(j)\}d(j,t_c)/2$.

[41] The ratio of the highest to lowest decile is just over two; the ratio of the highest quartile to the lowest quartile is just under two.

In Table 8, we report the parameter estimates that minimize the root mean squared error between the model prediction and the data for the change in low-skill share between 1881 and 1971. We also report bootstrapped standard errors calculated from grid search estimates on 1,000 resamples. The main parameter commanding the return to the mean, $\delta = 0.08$, implies that half of the gap between neighborhoods would be bridged after only 9 years. However, the model also estimates the presence of neighborhood effects counteracting the reversion process. The coefficient $\phi_1^e = 0.11$ capturing the continuous neighborhood effect is positive and the exponent $\phi_2^e = 0.89$ is less than one. While the continuous neighborhood effect is positive, it is, outwith the tail effect, too small to generate persistent sorting that is greater than that initially caused by the pollution. The estimated model finally captures the existence of a tail effect. The coefficient ϕ_1^b on the tail component is positive and the tail threshold is 0.76. This implies that the tail effect starts operating once a neighborhood is 26 percentage points higher in low-skill share than the city average. Finding a value of the exponent ϕ_2^b that is greater than one suggests that the tail costs are convex. In general, the tail effect is stronger than the reversion process and there is no return to the mean—once a neighborhood has suffered from enough pollution to cross this threshold, its long-run outcome in terms of skill-share is very similar than that originally caused by the pollution.

Model fit and over-identification checks To assess the validity of our model, we simulate the model using parameters estimated from 1881–1971 data and consider its performance in explaining persistence over the period 1971–2011. We use two statistics based on our observed 5,538 neighborhoods. First, we calculate the difference between the average low-skill share in areas with above and below within-city pollution in 1881. This measure is the average spread of low-skill share between the "East" and the "West",

$$\text{spread}(t) = E\left[S^l(j,t)\middle|p(j) > 0\right] - E\left[S^l(j,t)\middle|p(j) < 0\right] \tag{10}$$

where $p(j)$ is the city-normalized pollution level estimated for neighborhood j in 1881. With no persistent effect, this spread is zero. The second statistic is the correlation between low-skill shares in $t_1 = 1971$ and $t_2 = 2011$,

$$\rho_{t_2,t_1} = \frac{\sum_j \left(S^l(j,t_2) - \gamma\right)\left(S^l(j,t_1) - \gamma\right)}{\sum_j \left(S^l(j,t_1) - \gamma\right)^2} \tag{11}$$

The first two columns of Table 9 report results on these statistics in the model and the data. The model performs well in matching the targeted spread of low-skill occupations in 1971. More interestingly, the model matches quite well the spread in 2011 and the correlation between 1971 and 2011 despite the parameters being estimated to fit data for 1881–1971. We illustrate the model fit between 1971 and 2011 in Appendix Figure A11.

Note, however, that the 2011 spread of low-skill occupations is slightly lower in the model than in the data, and so is the measure of persistence over 1971–2011. Indeed, the model does not account for any policy shifts that occurred after 1971. One important policy after 1979 relates to Thatcher's reform of social housing. We extend the model to incorporate this.

Social housing We model social housing as a disamenity[42] and as being occupied by low-skill occupations. Up to 1979, we assume—as observed in the data—that social housing is orthogonal to past atmospheric pollution. Since all neighborhoods are equally affected, there are no consequences for land values and the distribution of low-skill occupations. Once social housing is liberalized, however, that part of the housing stock can enter the free market. Workers who are initially located in areas with better (worse) amenities can now ask for high (low) prices for their properties. The distribution of social housing thus converges to the same distribution as that of low-skill occupations (a process which is, in the data, completed by 1991).

The third column of Table 9 reports the model output with social housing liberalization ('SH-L' in the Table) against the baseline model and the data. The model fit for 1971–2011, either captured by the 2011 spread or the correlation between 1971 and 2011, substantially improves: the liberalization of social housing caused greater persistence in the distribution of deprivation because it removed a random component (the location of social housing) which was bringing neighborhoods closer to the city average. Our estimates suggest that about 20% of the remaining gradient between polluted and spared neighborhoods can be attributed to this reform.[43]

Counterfactual experiments We now provide two sets of counterfactual experiments to understand the role of non-linearities in the dynamics of segregation.

In a first set of experiments, we use the baseline model and impose a hypothetical construction boom in social housing in 1979, increasing the social housing stock from

[42]We select a coefficient on the social housing share to target the spread in 2011.

[43]While the original intent of Thatcher's policy was to reduce inequality by providing a route for working class households to step on the housing ladder, its consequence appears to have been to lengthen the shadow of the Industrial Revolution and set back the slow decay of neighborhood sorting.

30% to 40% or 45%. As can be seen in columns 2 and 3 of Table 10, even a substantial investment in social housing would have been ineffective in reducing the persistence of segregation over the period. With our estimated neighborhood effects, social housing programs would appear to be a costly means of reducing spatial inequalities. This result comes from the fact that there are not many neighborhoods that are just above the tail threshold, and few of them would revert to the city average even with a more uniform distribution of low-skilled workers, as implied by the social housing expansion. This intuition also holds in the next set of experiments.

In a second set of experiments, we vary the initial pollution exposure for all neighborhoods by ±25% to explore the quantitative impact of the pollution disamenity on the subsequent persistence of spatial inequalities. As can be seen in columns 4 and 5 of Table 10, higher (lower) initial pollution increases (decreases) the spread of low-skill occupations across a city in 1971. More importantly, the initial distribution of the pollution disamenity plays a role in the subsequent dynamics of persistence: a 25% higher exposure to pollution markedly increases the correlation between 1971 and 2011 (0.59 against 0.40) while a symmetric 25% lower exposure to pollution has little effect. This result is driven by the tail behavior in the underlying persistence mechanism and the number of neighborhoods on each side of the tipping threshold.

We can exploit variation across cities in the data to provide an over-identification test for this last quantitative prediction. Cities in our sample have similar shares of low-skilled workers but they widely differ in exposure to pollution. We define the city gradient in pollution as the difference between the 10% and 90% percentiles and divide cities in two groups of equal size: those with above-median pollution gradient and cities with below-median pollution gradient. We display in Appendix Figure A12 the dynamics of persistence in the two sets of cities, and the distribution of pollution within these cities. As apparent in the right panels, there is reversion to the mean in cities with a low share of polluted areas. The large correlation in 1971 steadily drops and becomes non-different from 0 in 2011. By contrast, cities with a high pollution gradient (left panels) do not experience any reversion to the mean.[44]

In the following section, we provide a more granular investigation of factors driving the dynamics of persistence. First, we look at the unexpected need to redevelop inner-city neighborghoods that were damaged during the "The Blitz" of 1940–1941, and we exploit the quasi-random location of bomb damage. We then provide indirect evidence of local neighborhood tipping and we show how persistence in occupational structure for a "cell" may depend on its neighbors.

[44]This observation has consequences at the city level. We show in Appendix Figure A13 that cities with high within-city dispersion in past pollution are more segregated nowadays.

6.3 Local factors driving the dynamics of persistence

A shortcoming of the previous 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 A14), and define a dummy *Bombs* equal to 1 if at least one bomb impact has been recorded within the LSOA between 1940 and 1945.[45] We then analyze bomb damages as shock to neighborhoods that induced changes in the local occupational structure.

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:

$$Y_{it} = \sum_{\tau=1971}^{2011} \beta_\tau P_i \times \mathbb{1}_{\tau=t} + \sum_{\tau=1971}^{2011} \beta^b_\tau P_i \times \mathbb{1}_{\tau=t} \times Bombs_i + \sum_{\tau=1971}^{2011} \gamma_\tau \mathbf{X}_i \times \mathbb{1}_{\tau=t} + \nu_i + \delta_t + \varepsilon_{it} \quad \text{(S3)}$$

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 2. 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^b_τ captures the difference between bombed and spared LSOAs in this relative long-term impact of pollution.

We report the estimates of β^b_τ in Panel A of Table 11. 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. Figure 7 illustrates the differential long-term effects of pollution, β^b_τ and $\beta_\tau + \beta^b_\tau$, 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.

We consider these findings as possibly informative about the impact of urban renewal policies and slum clearance.[46] Indeed, bomb damage may be considered

[45] The list of cities and additional details about data sources can be found in Appendix C. We omit Greater London from the analysis as we were unable to ensure that our scans were exhaustive.

[46] Slum clearances in England were numerous during the Interwar period, following the Housing Act of 1930 and the Housing Act of 1936. However, it is not possible to extract exogenous variation

exogenous at the local level once controlled for distance to industries. Following 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).[47] 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 Schelling (1971)'s theory of neighborhood tipping. 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, 2016; Couture and Handbury, 2017).

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 neighbours 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 S3 with the dummy *Homophily* instead of the dummy *Bombs*, and we report the results in Panel B of Table 11 and Figure 7. 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 7 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 almost 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

as they were explicitly targeting deprived but well-located neighborhoods.

[47]Note, however, that we run specification S3 for the normalized share of social housing in Panel A of Table 11, and we find that social housing is less correlated with historical pollution in bombed neighborhoods.

at once, as in Barcelona or London before their respective Olympic Games.

7 Conclusion

This paper presents a plausible explanation for the 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. A very unequal distribution of pollution exposure induced a sorting process which left the middle and upper class in the relatively less polluted neighborhoods. Our empirical analysis relies on precise pollution estimates and identifies neighborhood sorting at a highly local level: the illustrative East/West gradient reflects a global 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 and data around the peak time of coal use in 1881 to show that rising pollution set off the assumed 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 the high- and low-income equilibrium over time. We use a simple quantitative model to estimate the structure of neighborhood effects. Our estimates imply large non-linear effects with tipping-like behavior, and we replicate quite well the subsequent dynamics between 1971 and 2011. We then supplement the model and evaluate how urban redevelopment due to bomb damage and spatial correlation in occupational structure affect the dynamics of persistence.

Our findings hold at least two important implications. First, the success of urban policies to revitalize deprived areas may depend on their position relative to the tipping point, and the spatial distribution of poverty. As outlined by our findings, there are non-linearities in neighborhood effects and very deprived neighborhoods would need a large push to reach the tipping point. This observation leads to a second implication for countries like China where pollution currently presents a major challenge. Beside the well documented short-run effects of pollution exposure on health, there are long-run consequences of an uneven pollution exposure across space: pollution may induce large spatial inequalities that survive de-industrialization.

References

Anderson, Michael L., "As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality," Working Paper 21578, National Bureau of Economic Research September 2015.

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," Technical Report 19881 February 2014.

Baum-Snow, Nathaniel and Daniel A Hartley, "Accounting for central neighborhood change, 1980–2010," Technical Report 2016.

Beach, Brian and W. Walker Hanlon, "Coal Smoke and Mortality in an Early Industrial Economy," Working Paper 2016.

Benabou, Roland, "Workings of a City: Location, Education, and Production," *Quarterly Journal of Economics*, 1993, *108* (3), 619–52.

Brooks, Leah and Byron Lutz, "Vestiges of Transit: Urban Persistence at a Micro Scale," Technical Report June 2016.

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.

Clay, Karen and Werner Troesken, "Did Frederick Brodie Discover the World's First Environmental Kuznets Curve? Coal Smoke and the Rise and Fall of the London Fog," in Gary D. Libecap and Richard H. Steckel, eds., *The Economics of Climate Change: Adaptations Past and Present*, University of Chicago Press, 2011, pp. 281–309.

_ , 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.

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

Couture, Victor and Jessie Handbury, "Urban Revival in America, 2000 to 2010," Working Paper, National Bureau of Economic Research 2017.

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 I. Dingel Jonathan, "The Comparative Advantage of Cities," *NBER Working Paper*, 2014, (20602).

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," Working Paper 22796, National Bureau of Economic Research November 2016.

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," Working Paper 24157, National Bureau of Economic Research December 2017.

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," Working Paper 2016.

Herrnstadt, Evan and Erich Muehlegger, "Air Pollution and Criminal Activity: Evidence from Chicago Microdata," Working Paper 2015.

Heyes, Anthony and Soodeh Saberian, "Air Pollution Causes Violent Crime," Working Paper 2015.

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.

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 Impact of Short Term Exposure to Ambient Air Pollution on Cognitive Performance and Human Capital Formation,” Working Paper 20648, National Bureau of Economic Research October 2014.

Lee, Sanghoon and Jeffrey Lin, “Natural Amenities, Neighborhood Dynamics, and Persistence in the Spatial Distribution of Income,” *The Review of Economic Studies*, 2017, *forthcoming.*

Logan, Trevon and John M. Parman, “The National Rise in Residential Segregation (with Trevon Logan), forthcoming, Journal of Economic History,” *Journal of Economic History*, 2016, *forthcoming.*

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.

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

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

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," 2007.

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

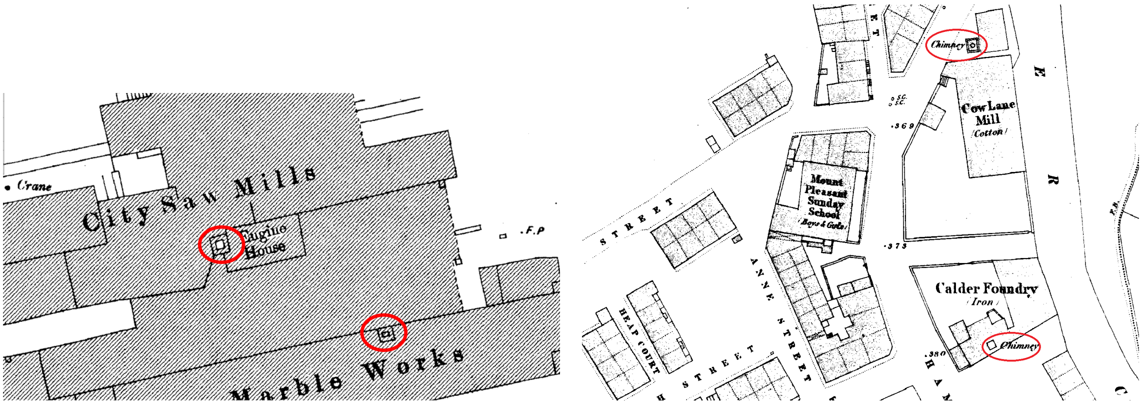
Williamson, Jeffrey G., "Earnings Inequality in Nineteenth-Century Britain," *The Journal of Economic History*, 1980, *40* (3), 457–475.

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

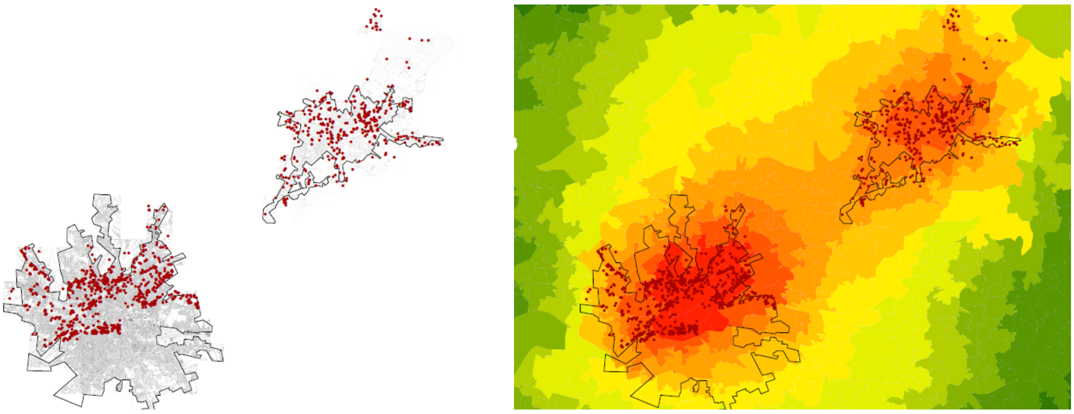
Figure 1. Ordnance Survey maps—chimney symbols.



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

Sources: Ordnance Survey Maps—25 inch to the mile, 1842–1952. Four different symbols for chimneys are circled.

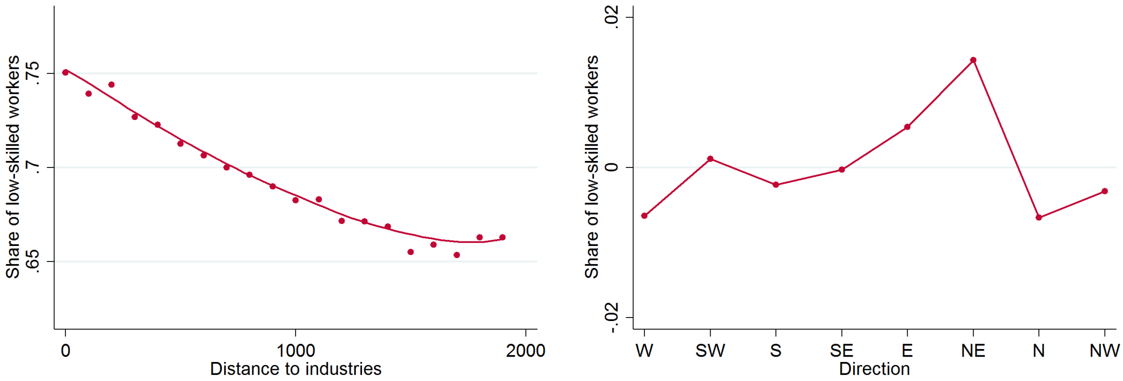
Figure 2. Aggregating pollution sources (Manchester).



(a) Locating chimneys. (b) Pollution imprints.

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 red dots.

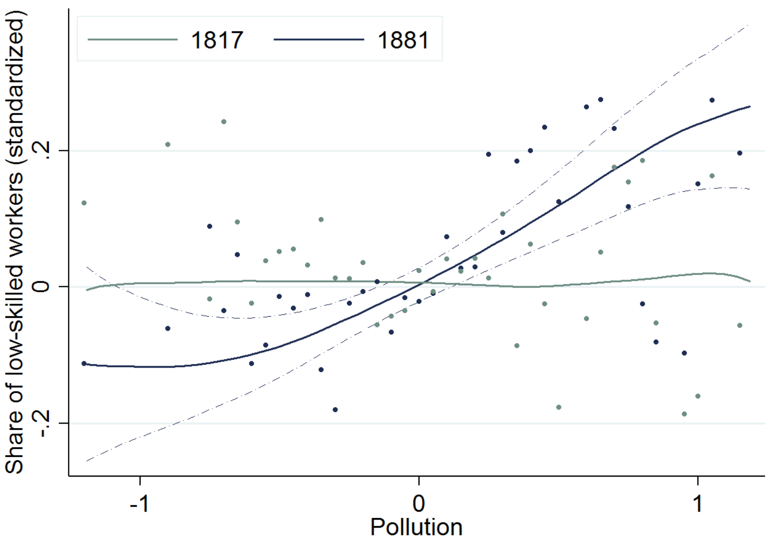
Figure 3. Shares of low-skilled workers and position relatively to pollution sources.



(a) Distance. (b) Direction.

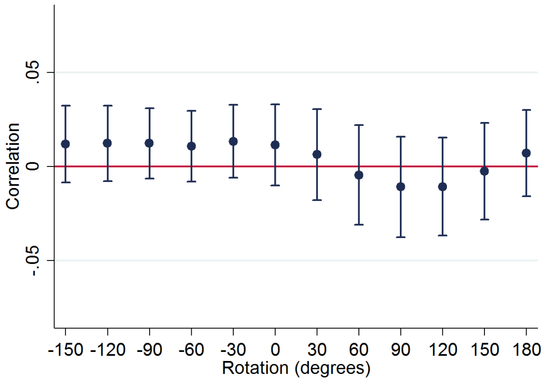
Notes: This Figure represents the relationship between the share of low-skilled workers in 1881 and the relative position to a pollution source. In this exercise, the unit of observation is a couple neighborhood×chimney where a neighborhood is a Census cluster of households with the same geolocation in 1881 (about 150 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 (park, church, school, university etc.), and averaged over 8 main directions with respect to the pollution source. *NE* stands for North-East, indicating that the household is located toward the North-East direction, from the standpoint of the pollution source.

Figure 4. Pollution (x-axis) across neighborhoods and shares of low-skilled workers (y-axis) in 1817 and 1881.

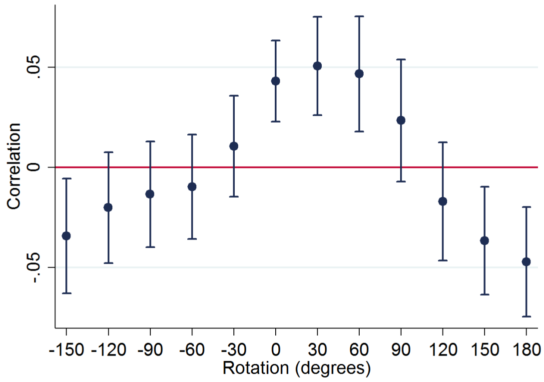


Notes: This Figure represents the relationship between the (standardized) shares of low-skilled workers in 1817 (teal) and 1881 (blue) and our (standardized) measure of past pollution. We consider the residuals of all measures once cleaned by the *topography* controls, the *amenities* controls and the *lat./lon.* controls (see Table 2). We create 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).

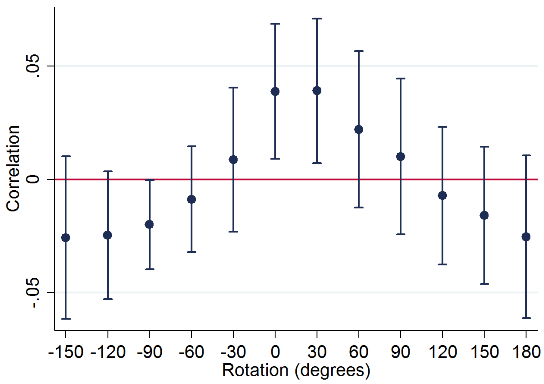
Figure 5. Rotating wind patterns in 1817 and 1881.



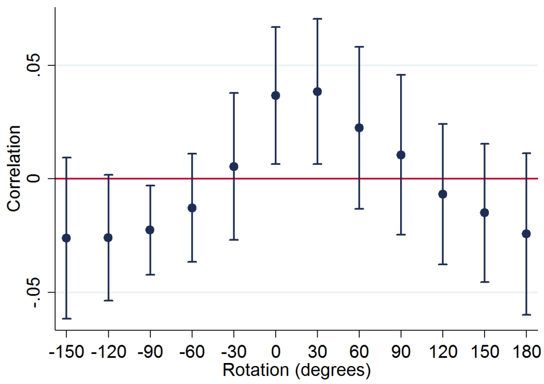
(a) 1817.



(b) 1881.



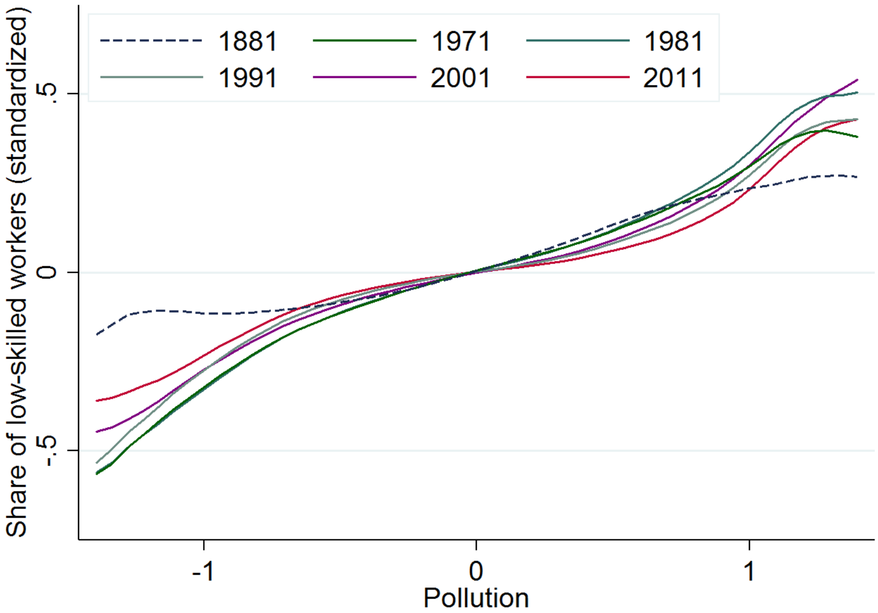
(c) 1881 (parish FEs).



(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 2, Column 6, and the measure of *Static pollution* capturing proximity to the pollution source (see Table 4). 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.

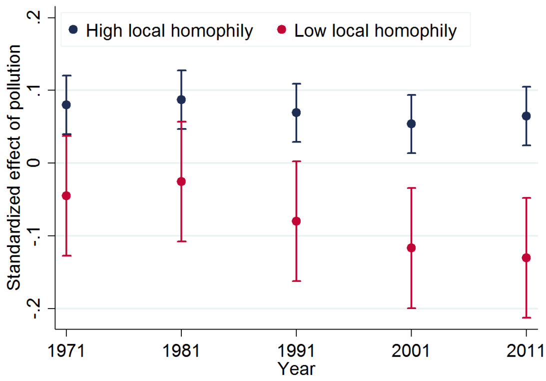
Figure 6. Pollution (x-axis) across neighborhoods and shares of low-skilled workers (y-axis) in 1881, 1971, 1981, 1991, 2001 and 2011.



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 7. 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. The sample is the baseline sample of LSOAs outside Greater London, and 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. See Section 6 for a detailed description of both specifications.

Table 1. Descriptive statistics and variance decomposition.

		Pollution		Standard deviation		
VARIABLES	Mean	high	low	total	between	within
		Air pollution				
Normalized pollution	-.034	.233	-.298	.928	.596	.560
		Population measures				
*1817**						
Low-skilled workers	.785	.788	.780	.110	.090	.072
High-skilled workers	.093	.083	.102	.086	.085	.068
Farmers	.122	.128	.117	.091	.065	.047
Property tax (log)	9.93	10.04	9.82	1.25	1.04	.076
1881						
Low-skilled workers	.600	.627	.574	.255	.149	.235
High-skilled workers	.278	.287	.268	.241	.125	.216
Farmers	.121	.085	.157	.199	.175	.180
2011						
Low-skilled workers	.587	.604	.569	.173	.117	.123
High-skilled workers	.413	.395	.430	.173	.117	.123
		Topography controls				
Maximum elevation (m)	72.8	66.2	79.3	66.0	63.2	31.8
Minimum elevation (m)	52.5	49.9	55.2	48.9	44.1	19.6
Mean elevation (m)	62.3	57.8	66.7	55.8	51.6	23.4
Distance canals (km)	6.12	5.47	6.76	14.9	19.1	1.30
		Amenities controls				
Distance town hall (km)	4.64	4.10	5.18	5.35	4.72	1.27
Distance parks (km)	9.57	9.37	9.77	23.9	28.8	1.15
Share LSOA within city	.296	.404	.190	.417	.245	.296
Area (square km)	.939	.599	1.27	3.99	7.99	3.49
Distance heavy (km)	2.48	1.92	3.03	6.36	8.81	.972
Distance light (km)	5.31	4.93	5.69	13.5	17.6	1.15
		Direction				
Easting (km)	453	454	452	74.9	76.9	1.90
Northing (km)	291	291	292	123	126	1.92

Notes: All statistics are computed using the baseline sample of 5,538 LSOAs. Standard deviations are decomposed into between- and within-city standard deviations. * Shares in 1817 are computed at the parish-level, which explains the lower variance. Latitude and longitude are reported as the eastings and northings in the Ordnance Survey National Grid. The samples of high- and low-within-city pollution are defined with respect to the median city pollution.

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	.0377	.0354	.0337
	(.0071)	(.0072)	(.0068)	(.0066)	(.0067)	(.0069)
	[.1737]	[.1689]	[.1614]	[.1491]	[.1397]	[.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 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.

Table 4. Pollution and shares of low-skilled workers in 1881—counterfactual pollution imprints and residential pollution.

Share of low-skilled workers	(1)	(2)	(3)	(4)
Pollution	.0442	.0355	.0427	.0377
	(.0082)	(.0113)	(.0139)	(.0137)
	[.1746]	[.1403]	[.1684]	[.1487]
Mirror Pollution	-.0120	-.0260	-.0273	-.0262
	(.0068)	(.0145)	(.0146)	(.0145)
	[-.0474]	[-.1027]	[-.1080]	[-.1035]
Static Pollution		.0209	.0253	.0228
		(.0197)	(.0210)	(.0209)
		[.0828]	[.0831]	[.0900]
Placebo Industry			-.0136	-.0114
			(.0153)	(.0152)
			[-.0536]	[-.0452]
Residential Pollution				.0138
				(.0046)
				[.0546]
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. 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. *Mirror Pollution* (resp. *Static Pollution*) is the counterfactual pollution exposure using a wind rose rotated 180 degrees (resp. 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 3 for additional details about the construction of these variables.

Table 5. Pollution and shares of low-skilled workers in 1881—2SLS specification.

First stage	Pollution			
	(1)	(2)	(3)	(4)
Pollution (waterways)	.2020	.1834		
	(.0346)	(.0384)		
Pollution (steam engines)			.1700	.1937
			(.0325)	(.0399)
Second stage	Share of low-skilled workers (1881)			
	(1)	(2)	(3)	(4)
Pollution	.0936	.0695	.1011	.0801
	(.0296)	(.0358)	(.0379)	(.0402)
	[.3695]	[.2743]	[.3991]	[.3160]
Observations	4,830	4,557	2,519	1,860
F-statistic	33.91	22.81	28.00	23.49
OLS coefficient	.0216	.0208	.0221	.0193
Sample	Canal>250m	Canal>500m	Mill<10km	Mill<5km
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. 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). The variable *Pollution (steam engines)* is the second 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 respectively, we exclude LSOAs within 500 meters and 250 meters of a waterway. In columns 3 and 4 respectively, we exclude LSOAs outside of a range of 10 and 5 kilometers around a textile factory.

Table 6. Pollution and shares of low-skilled workers in 1971, 1981, 1991, 2001 and 2011.

Share of low-skilled workers	1971	1981	1991	2001	2011
Pollution	.0243	.0309	.0388	.0374	.0354
	(.0046)	(.0050)	(.0063)	(.0063)	(.0057)
	[.1914]	[.2203]	[.2071]	[.2298]	[.2028]
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 7. Selected parameters (see Section 6).

Parameter		Value	Rationale
θ^h	High income	2	Williamson (1980), highest quartile to the lowest
θ^l	Low income	1	Williamson (1980)
$\tilde{\gamma}$	Low-skill share	0.50	Normalization
α	Pollution sensitivity	0.102	Correlation pollution/occupation in 1881
d	Initial amenity	1	Normalization

Note: The sensitivity α of low-skill share to pollution is calibrated using the correlation between within-city residuals in low-skills and atmospheric pollution in 1881.

Table 8. Estimated parameters (see Section 6).

Parameter	Description	Estimate	Standard error
ϕ_1^e	Coefficient for the continuous effect	0.11	0.04
ϕ_2^e	Curvature for the continuous effect	0.89	0.08
ϕ_1^b	Coefficient for the tail effect	0.10	0.06
ϕ_2^b	Curvature for the tail effect	1.45	0.30
$\bar{S}$	Tail point	0.76	0.08
δ	Depreciation factor	0.08	0.03

Note: The initial grid search is over the following ranges: $\phi_1^e = [0, 0.3]$; $\phi_2^e = [0, 1.5]$; $\phi_1^b = [0, 0.3]$; $\phi_2^b = [0, 2.5]$; $\bar{S} = [0.50, 0.90]$; $\delta = [0, 0.15]$. Bootstrapped standard errors are calculated from grid searches on 1,000 random resamples with replacement.

Table 9. Baseline model and model with social housing liberalization against data.

	Data	Baseline	SH-L
Spread in 1971	.0550	.0555	.0555
Spread in 2011	.0278	.0235	.0281
Correlation $\rho_{2011,1971}$	.4337	.4010	.4331

Note: SH-L is the baseline model augmented by the social housing liberalization of the Thatcher government in 1979 (See Section 6).

Table 10. Counterfactual experiments (alternative social housing and pollution exposure).

		Social housing		Pollution	
	Baseline	SH-40	SH-45	-25%	$+25\%$
Spread in 1971	.0555	.0555	.0555	.0421	.0737
Spread in 2011	.0235	.0229	.0218	.0183	.0358
Correlation $\rho_{2011,1971}$	.4010	.3698	.3167	.3885	.5944

Note: In columns 2 and 3, SH-N are experiments where we introduce a $N\%$ ($N = 40, 45$) social housing supply at all locations. In columns 4 and 5, we vary the initial pollution estimates for all neighborhoods by $\pm 25\%$.

Table 11. Persistence of neighborhood sorting—role of bombings and local homophily.

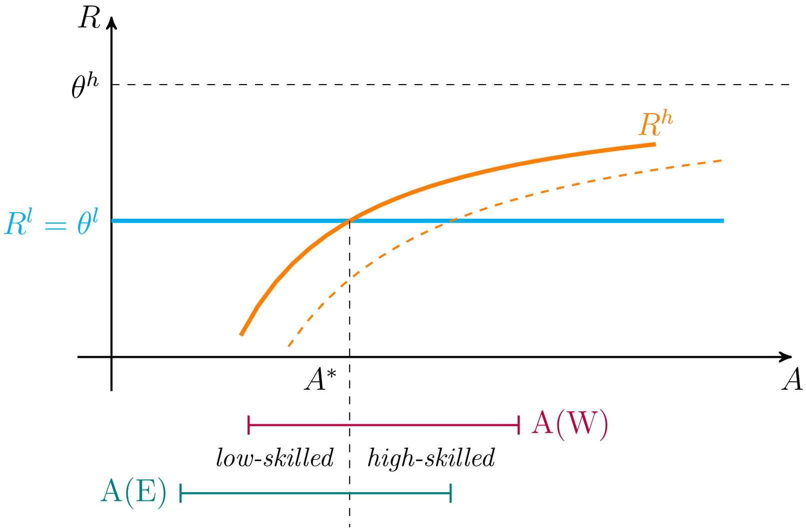
Panel A: role of bombing		
	Share of low-skilled	Share of social housing
Pollution × Bombs × 1971	-.1870	
	(.0432)	
Pollution × Bombs × 1981	-.1719	-.0273
	(.0432)	(.0255)
Pollution × Bombs × 1991	-.1932	-.0619
	(.0432)	(.0255)
Pollution × Bombs × 2001	-.1112	-.0742
	(.0432)	(.0255)
Pollution × Bombs × 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, the sample is the baseline sample of LSOAs outside Greater London, and *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 neighbours within the same MSOA lie on the same side of the city average. See Section 6 for a detailed description of both specifications.

ONLINE APPENDIX—not for publication

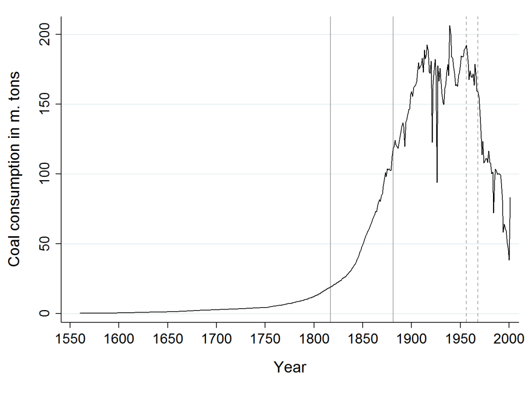
A Additional figures and tables

Figure A1. Pollution and neighborhood sorting: equilibrium with $\gamma = \frac{1}{2}$.

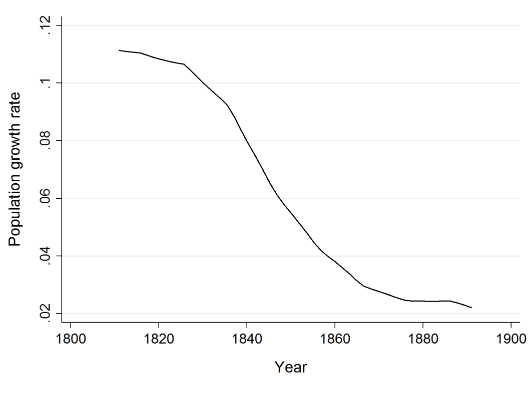


Notes: the x-axis represents the level of consumptive amenities and the y-axis is the rent. $A(W)$ and $A(E)$ depict the distribution of amenities in the two neighborhoods $\{E, W\}$. We assume that the amenity levels overlap across neighborhoods: a high income agent prefers the nicest location in the polluted East to the worst location in the non-polluted West.

Figure A2. Migration and coal consumption during the Industrial Revolution.



(a) Coal consumption (1560–2001)



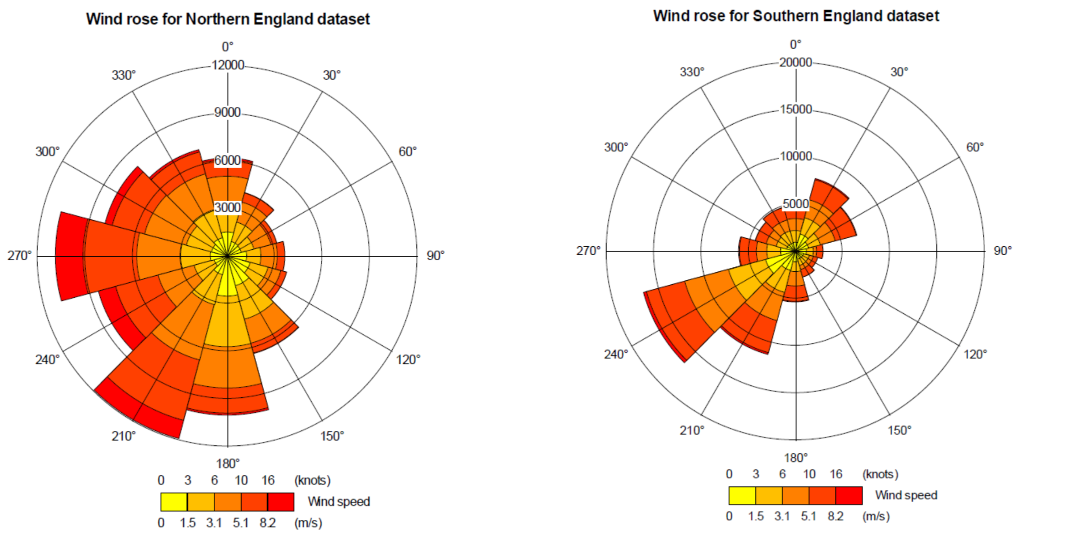
(b) Urban growth rates (1801–1891)

Notes: The left panel illustrates the increase in coal consumption over the period 1560–2001. 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 grey lines indicate the years 1817 and 1881, while the dashed grey 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 our sample cities.

Figure A3. 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.

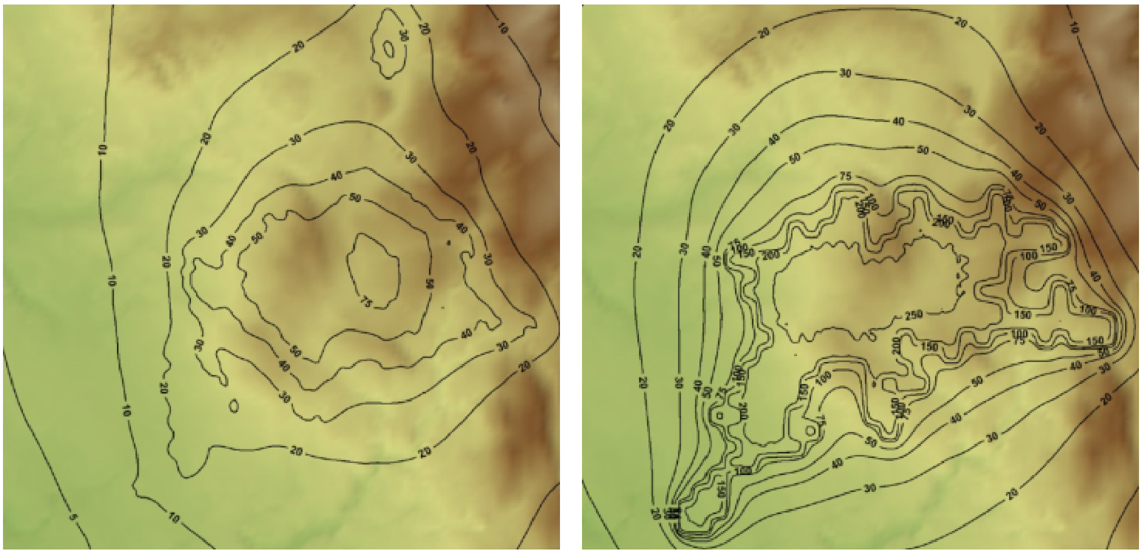
Figure A4. Wind roses differences across two sets of meteorological conditions.



(a) North England. (b) South England.

Sources: Met Office—10-year statistical meteorological data. We use 5 different sets of meteorological conditions across England.

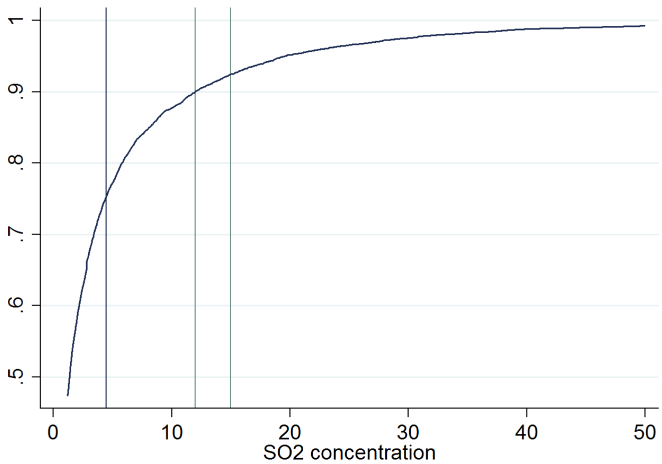
Figure A5. Topography and historical air pollution—the example of Oldham.



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

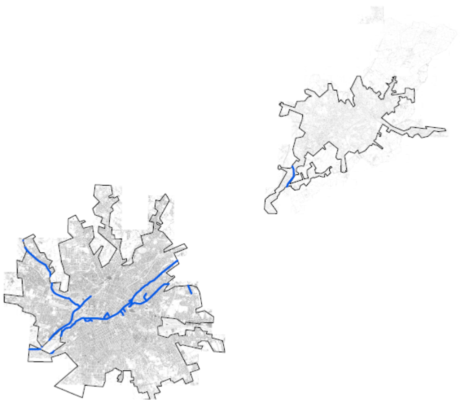
Sources: ADMS 5. These maps shows elevation (from green to brown) and the level lines for industrial and residential pollution in Oldham.

Figure A6. Cumulative of pollution in our baseline sample of 5,538 LSOAs and National Ambient Air Quality Standards (12-15 $\mu g/m^3$).



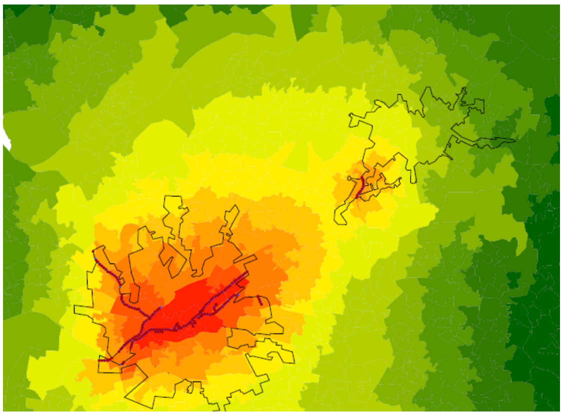
Sources: This Figure represents the cumulative distribution of SO2 concentration as predicted by the distribution of pollution sources, the emission intensity and air pollutant dispersion. See Section 3 and Appendix Section C for additional details.

Figure A7. An illustration of the 2SLS empirical approach in Manchester—hypothetical chimneys located along 1827 waterways.

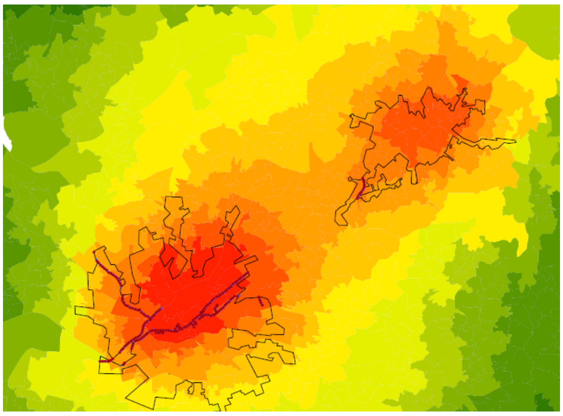


(a) City Map

(b) Locating hypothetical chimneys along waterways.



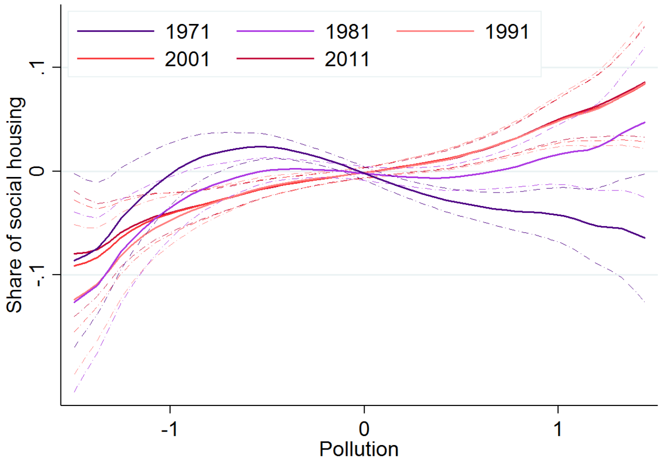
(c) Predicted pollution instrument.



(d) Simulated historical pollution.

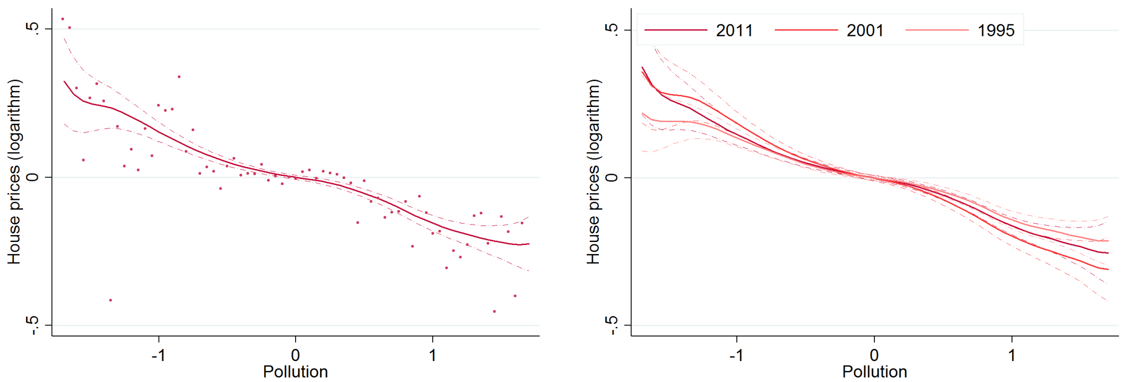
Sources: Authors' calculations using Ordnance Survey Maps—25 inch to the mile, 1842–1952 and the ADMS 5 Air Pollution Model. Chimneys are indicated with a red dot, and 1827 natural waterways with blue lines.

Figure A8. 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, geographic and topographic controls.

Figure A9. House transaction prices (y-axis) and pollution (x-axis) across neighborhoods—average and evolution between 1995 and 2011.

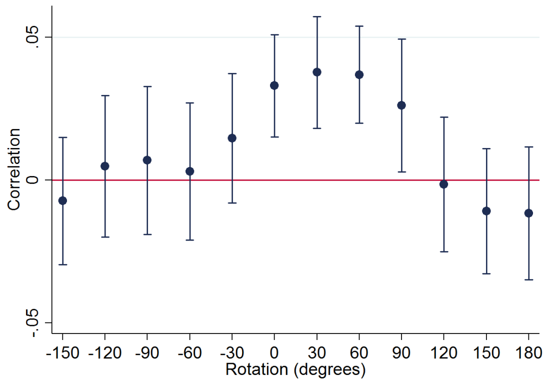


(a) Average

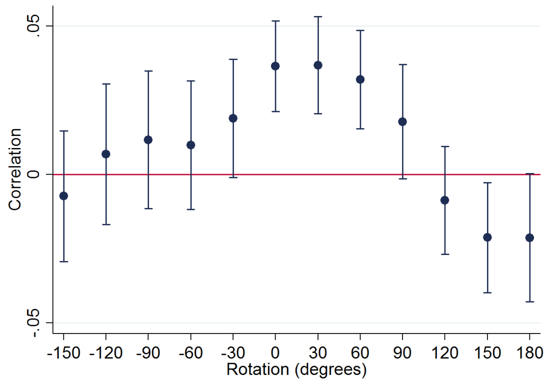
(b) Evolution

Notes: The left (resp. right) panel represents the relationship between the (logarithm of the) average transaction prices between 2000 and 2011 (resp. in 1995, 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. For the sake of exposure, 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 A10. Rotating wind patterns in 1991 and 2011.



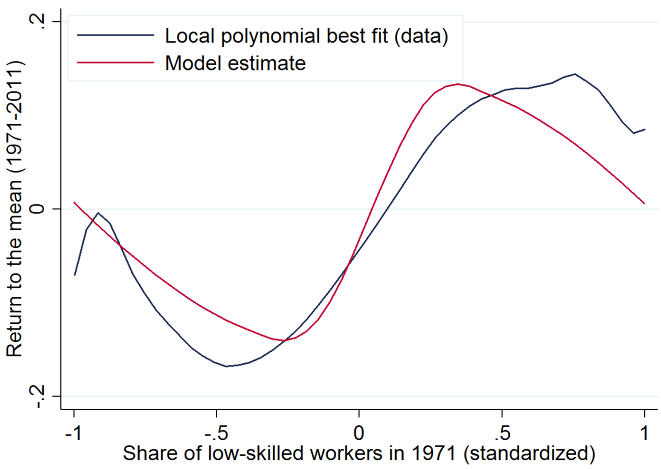
(a) 1991



(b) 2011

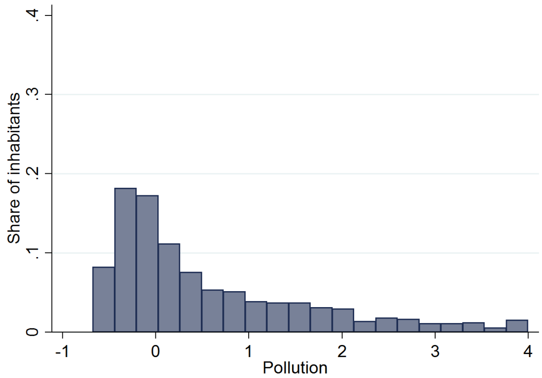
Notes: These Figures represent the conditional correlations between the shares of low-skilled workers (in 1991 and in 2011) and counterfactual measures of past pollution rotated in steps of 30 degrees. Each dot represents the estimate in a specification including the controls reported in Table 2, Column 6, and the measure of *Static pollution* capturing proximity to the pollution source (see Table 4). Standard errors are clustered at the parish-level, and the lines represent 5% confidence intervals.

Figure A11. Model and data estimates for the persistence between 1971 and 2011.

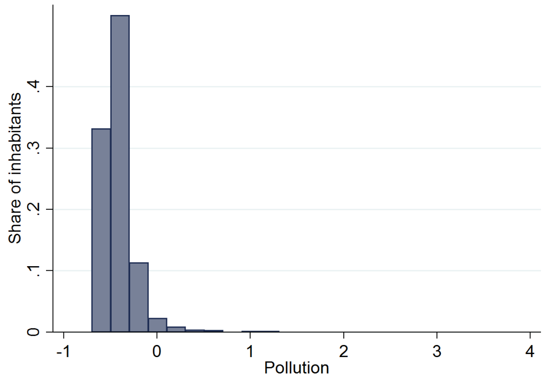


Notes: The Figure represents (i) the locally weighted polynomial regressions on all observations between the residual shares of low-skilled workers in 1971 (standardized between -1 and 1) and a measure capturing the return to the mean between 1971 and 2011, and (ii) the model estimates for the same relationship. The *return to the mean* measure is the difference between the residual shares of low-skilled workers in 1971 and 2011. If the process was an AR(1) process of parameter θ, the graph would depict a line of slope $1-\theta$. Accordingly, for an initial residual share of 0.5 in 1971, about 0.15 is subtracted from the 2011 measures corresponding to $1-\theta = .3$. At both ends, there is no return to the mean. The residuals of all measures are cleaned of city fixed effects, and *geography* and *topography* controls.

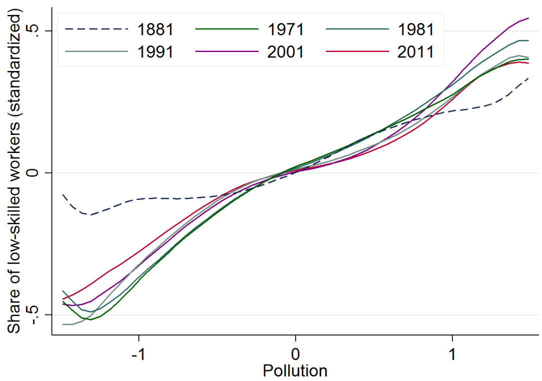
Figure A12. Importance of the city-wide distribution of pollution in sorting across neighborhoods.



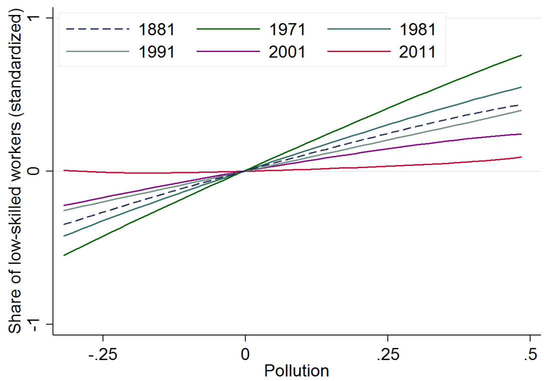
(a) Pollution (high gradient).



(b) Pollution (low gradient).



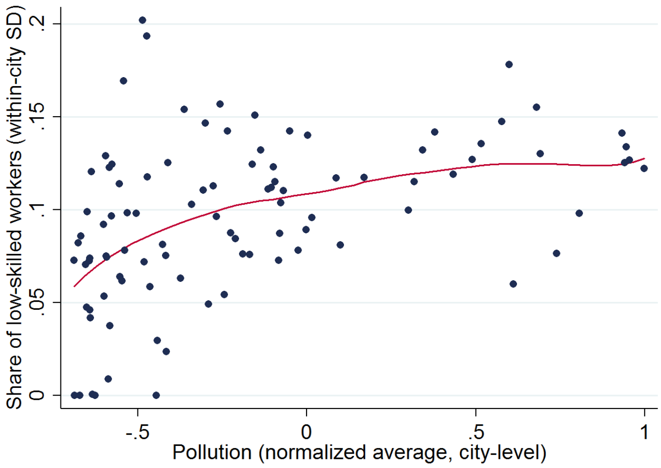
(c) Persistence (high gradient).



(d) Persistence (low gradient).

Notes: This Figure represents the relationship between the shares of low-skilled workers and our (standardized) measure of past pollution in two sets of cities. In the left panels (resp. right panels), we keep cities for which the pollution gradient between the 10% and 90% percentiles is above (resp. below) the median. We consider the residuals of all measures once cleaned by city fixed effects, *geography* and *topography* controls. For the sake of exposure, we group neighborhoods, create 100 bins of neighborhoods with similar past pollution and represent the average shares of low-skilled workers within a pollution-bin. The lines are locally weighted regressions on the observed sample.

Figure A13. Within-city dispersion in past pollution (x-axis) and city-level segregation in 2011 (y-axis).



Notes: This Figure represents the relationship between the within-city standard deviation in the share of low-skilled workers (across LSOAs), and the dispersion in past pollution as measured by the difference between the 10% and 90% percentile within city.

Figure A14. Bomb Census maps—an example in Birmingham chimneys.

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

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

Industry	Average coal use C_i $m.tons/year$	Average weight E_i
Brewery	19.4	0.36
Bricks	48.9	1.05
Chemical	40.1	0.84
Food processing	12.0	0.77
Metal	43.7	1.00
Mining	28.9	3.55
Paper	9.7	2.47
Shipbuilding	6.1	1.02
Tannery	12.1	0.17
Textile production	10.1	0.47
Wood	5.4	0.10
Works	-	0.10

Source: Hanlon (2016) 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 industry "Metal and engine manufacturing". The measure E_i is set to the minimum average value for chimneys classified as "Works".

Table A2. Air Pollution measures in the neighborhoods of Manchester.

Station	Deposits $m.tons/m^2$	Model estimates $\mu g/m^3$
Ancoats hospital	30.59	119.95
Philips Park	22.59	74.49
Whitworth Street	22.51	102.47
Queen's Park	20.18	70.00
Moss Side	18.69	29.11
Whitefield	15.53	11.92
Fallowfield	13.24	17.69
Davyhulme	12.68	6.93
Cheadle	10.63	9.40
Bowdon	6.25	0.02

Source: First Annual Report of the Sanitary Committee on the Work of the Air Pollution Advisory Board, 1915.

Table A3. 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 A4. Pollution and shares of low-skilled workers in 1881—sensitivity analysis to fixed effects, clusters and sample selection.

Panel A: Fixed effects	Share of low-skilled workers (1881)		
	(1)	(2)	(3)
Pollution	.0387	.0377	.0393
	(.0091)	(.0115)	(.0118)
	[.1527]	[.1490]	[.1551]
Observations	5,538	5,538	5,538
Fixed effects	Parish	Ward	MSOA
Panel B: Clusters	Share of low-skilled workers (1881)		
	(1)	(2)	(3)
Pollution	.0337	.0337	.0337
	(.0058)	(.0066)	(.0086)
	[.1332]	[.1332]	[.1332]
Observations	5,538	5,538	5,538
Clusters	MSOA	Ward	City
Panel C: Sample	Share of low-skilled workers (1881)		
	(1)	(2)	(3)
Pollution	.0297	.0642	.0343
	(.0070)	(.0126)	(.0069)
	[.1172]	[.2536]	[.1353]
Observations	3,661	4,395	5,247
Excluding...	London	NW	NE

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. A *MSOA* (Medium Super Output Area) is the second smallest unit in the census, and there are 1,800 MSOAs in our sample. A *ward* is an electoral ward (election for local councils), and there are 1,300 wards 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.

Table A5. Pollution and shares of low-skilled workers in 1881—sensitivity to the exclusion of suburbs/rural LSOAs

Share of low-skilled	(1)	(2)	(3)	(4)
Pollution	.0339	.0350	.0361	.0425
	(.0068)	(.0075)	(.0085)	(.0093)
	[.1340]	[.1383]	[.1425]	[.1677]
Observations	5,122	3,521	2,467	1,583
Sample	Hall$<$10km	Hall$<$5km	LSOA $\cap$ city	LSOA $\cap$ city$>$0.5
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. The set of extended controls include all controls of column 6 in Table 2. In columns 1 and 2 respectively, we exclude LSOAs outside of a range of 10 and 5 kilometers around the townhall. In columns 3 and 4 respectively, we exclude LSOAs whose share of area within the 1890 city borders is equal to 0, and lower than 0.5.

Table A6. 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.

Table A7. 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.

Table A8. 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 3). 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.

Table A9. Pollution, house prices and transactions (Nationwide and Land registry, 2009–2013).

	Nationwide		Land registry			
	House prices		House prices		Transactions	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Pollution	-.1035	-.0852	-.1116	-.0642	-.0473	-.0878
	(.0168)	(.0121)	(.0161)	(.0120)	(.0233)	(.0230)
	[-.1684]	[-.1385]	[-.2030]	[-.1168]	[-.0892]	[-.1656]
Observations	5,226	5,226	5,538	5,538	5,538	5,538
Fixed effects (city)	Yes	Yes	Yes	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes	Yes	Yes	Yes
Controls (house ch.)	No	Yes	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 (logarithm of the) average house prices (from Nationwide in columns 1 and 2, and Land registry in columns 3 and 4) and number of transactions (Land registry) between 2009 and 2013. 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 columns 4 and 6, controls for house characteristics include the average shares of detached, semi-detached, terraced houses and new houses for all transactions.

Table A10. Pollution and shares of low-skilled workers—controlling for amenities, housing characteristics and crime and education indicators.

Panel A: Amenities	1881		2011	
Share of low-skilled workers	(1)	(2)	(3)	(4)
Pollution	.0304	.0296	.0373	.0328
	(.0069)	(.0069)	(.0062)	(.0057)
	[.1202]	[.1169]	[.2138]	[.1880]
Observations	4,694	4,694	4,694	4,694
Extended controls	Yes	Yes	Yes	Yes
Controls (amenities 1881)	Yes	Yes	Yes	Yes
Controls (amenities 2011)	No	Yes	No	Yes
Panel B: Housing characteristics		2011		
Share of low-skilled workers	(1)	(2)	(3)	(4)
Pollution	.0469	.0413	.0245	.0142
	(.0087)	(.0063)	(.0050)	(.0039)
	[.2689]	[.2363]	[.1402]	[.0813]
Observations	1,472	5,226	5,226	5,226
Sample	New housing	All	All	All
Extended controls	Yes	Yes	Yes	Yes
Controls (building age)	Yes	Yes	Yes	Yes
Controls (social housing)	No	No	No	Yes
Controls (house characteristics)	No	No	Yes	Yes
Panel C: Education and crime		2011		
Share of low-skilled workers	(1)	(2)	(3)	(4)
Pollution	.0319	.0210	.0354	.0259
	(.0045)	(.0043)	(.0057)	(.0044)
	[.1825]	[.1206]	[.2028]	[.1486]
Observations	5,538	5,538	5,538	5,538
Extended controls	Yes	Yes	Yes	Yes
Controls (school supply)	Yes	Yes	No	No
Controls (composition/scores)	No	Yes	No	No
Controls (police station)	No	No	Yes	Yes
Controls (crime)	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 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. Controls for amenities in 1881 include the number of parks, schools, theaters, museums, churches, hospitals per 100 inhabitants at the LSOA level. Controls for amenities in 2011 include the number of parks, schools, theaters, museums, churches, hospitals, public buildings (e.g., town halls), courts, police stations, bus or train stations, botanical gardens, banks and conference centers per 100 inhabitants at the LSOA level.

Table A11. Pollution and shares of low-skilled workers in 2011—counterfactual pollution patterns and residential pollution.

Share of low-skilled workers	(1)	(2)	(3)	(4)	(5)
Pollution	.0379	.0335	.0355	.0387	.0379
	(.0062)	(.0084)	(.0107)	(.0112)	(.0112)
	[.2172]	[.1918]	[.2034]	[.2216]	[.2173]
Mirror Pollution	-.0028	-.0100	-.0104	-.0111	-.0106
	(.0052)	(.0101)	(.0103)	(.0105)	(.0105)
	[-.0165]	[-.0573]	[-.0595]	[-.0637]	[-.0606]
Static Pollution		.0106	.0119	.0135	.0129
		(.0134)	(.0144)	(.0147)	(.0148)
		[.0612]	[.0682]	[.0773]	[.0740]
Placebo Industry			-.0038	-.0052	.0042
			(.0115)	(.0119)	(.0119)
			[-.0220]	[-.0298]	[-.0242]
Residential Pollution				-.0087	-.0091
				(.0046)	(.0045)
				[-.0503]	[-.0521]
Current Pollution					.0104
					(.0037)
					[.0598]
Observations	5,538	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. 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. *Mirror Pollution* (resp. *Static Pollution*) is the synthetic pollution exposure using a wind rose rotated 180 degrees (resp. 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 3 for additional details about the construction of these variables.

Table A12. Pollution and shares of low-skilled workers in 1971, 1981, 1991, 2001 and 2011—2SLS specification.

Share of low-skilled workers	1971	1981	1991	2001	2011
Pollution	.0300	.0406	.0341	.0461	.0494
	(.0164)	(.0198)	(.0257)	(.0193)	(.0213)
	[.2357]	[.2894]	[.1823]	[.2838]	[.2829]
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 column 1 of Table 5, 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.

B Proofs

Proof of Proposition 1. Recall that the total mass of land equals the total mass of workers equals 2.

Let $F(A)$ be the cumulative density of land with amenity level less than or equal to A within the city. Clearly, $F(A) = 0$ for $A < A^{min} \equiv \min_{j\in W,E}\left\{\min_{\ell\in\Omega(j)} A(j,\ell)\right\}$ and $F(A) = 2$ for $A > A^{max} \equiv \max_{j\in W,E}\left\{\max_{\ell\in\Omega(j)} A(j,\ell)\right\}$. Suppose that amenity levels across neighborhoods overlap in the sense that,

$$\max_{j\in W,E}\left\{\min_{\ell\in\Omega(j)} A(j,\ell)\right\} < \min_{j\in W,E}\left\{\max_{\ell\in\Omega(j)} A(j,\ell)\right\}.$$

Since amenities overlap, $F(A)$ is monotonically increasing and continuous in A over the interval $[A^{min}, A^{max}]$. As such, there is an $A^* \in [A^{min}, A^{max}]$ such that $F(A^*) = 2\gamma$. From Equation (4), landlords are indifferent to high- and low-skilled workers at A^* if high-skilled worker utility is $\bar{V}^{h*} = A^*(\theta^h - \theta^l)$. By (3) and (4), with $\bar{V}^{h*} = A^*(\theta^h - \theta^l)$ we have $R^h > R^l$ for all $A > A^*$ and $R^h \le R^l$ for all $A \le A^*$. Since $F(A^*) = 2\gamma$, the mass of land rented to low-skilled workers in the city satisfies (5).

The cumulative land density in the overlapping interval of amenities is,[48]

$$F(A) = \sum_{j\in\{W,E\}} A - \min_{\ell\in\Omega(j)} A(j,\ell),$$

$$\text{for} \quad A \in \left[\max_{j\in W,E}\left\{\min_{\ell\in\Omega(j)} A(j,\ell)\right\}, \min_{j\in W,E}\left\{\max_{\ell\in\Omega(j)} A(j,\ell)\right\}\right].$$

Suppose that the A^* such that $F(A^*) = 2\gamma$ is in this interval. The equilibrium is characterized by imperfect sorting in the sense that neither neighborhood fully specializes in high- or low-skilled workers. □

Proof of Lemma 1. With pollution emissions, in the overlapping interval of amenities we have,

$$F(A) = 2(A - d + \rho).$$

In equilibrium, where $F(A) = 2\gamma$ so $A^* = d + \gamma - \rho$. The share of low-skilled workers in each neighborhood is,

$$S^l(W) = \gamma - \eta\rho,$$
$$S^l(E) = \gamma + \eta\rho.$$

[48] The full expression is $F(A) = \sum_{j\in\{W,E\}} \frac{A - \min_{\ell\in\Omega(j)} A(j,\ell)}{\max_{\ell\in\Omega(j)} A(j,\ell) - \min_{\ell\in\Omega(j)} A(j,\ell)}$, but note that the denominator is equal to 1 by assumption that $x(\ell, j)$ is distributed uniformly over $[\phi(j), \phi(j) + 1]$.

Since $\eta\rho > 0$, the share of low-skilled workers in the West is less than the share in the East. Moreover, the greater the pollution intensity, ρ, or the stronger is the wind, η, the larger is the difference in the shares of low-skilled workers across neighborhoods. $\square$

Proof of Lemma 2. After $t = t_p$, pollution causes $\bar{\theta}(W,t) > \bar{\theta}$ by Lemma 1 and so accumulation of amenities by equation (7). The pollution, and consequent general amenities, may then cause $b(j,t)$ to accumulate by (8). That is, $d(W,t)$ may increase and $d(E,t)$ may decrease as a result of pollution. Let $d(t) = d(W,t) + d(E,t)$ be the spread of endogenous amenities. We can write for $t \geq t_c$, $F(A) = 2A - d(t)$, and,

$$A^* = \gamma + d(t)/2.$$

The share of low-skilled workers in each neighborhood is,

$$S^l(W,t) = \gamma - d(t)/2, \tag{12}$$

$$S^l(E,t) = \gamma + d(t)/2. \tag{13}$$

Since $d(j,t)$ amenities are persistent, $d(t) > 0$ permanently unless depreciation exists. Equations (12)-(13) then show that sorting persists even after $t = t_c$. $\square$

C Data sources

In this section, we describe 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 centres at the beginning of the nineteenth century, and select the nearest wave to the year 1890.[49] 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 digitisation.

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 recognisable 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.[50] In a second step, the mark is identified by a recognition algorithm as well as the associated identifier. The recognition algorithm

[49] 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).

[50] 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.

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 3) 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 A5.

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 digitised 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, birth place, 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 A14).

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 Table 11 and Figure 7) 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.[51]

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 Table 5.

The variable *Pollution (waterways)* of Table 5 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"

[51]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 A4) or to compute measures of neighborhood similarity (Table 11 and Figure 7).

House price data (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 Appendix Table A9, and we use housing type as dependent variable in Appendix Table A8 and as controls in Appendix Tables A9 and A10.

Building age data: We collect measures of the residential dwelling ages, grouped into three age bands, and collected by the Consumer Data Research Centre. We use these measures as dependent variables in Appendix Table A8 and as controls in Appendix Table A10.

Education data: 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,.[52]

Crime data: 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.[53] The number of crimes per inhabi-

[52]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.

[53]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.

tants in each category is used as dependent variable in Appendix Table A8.

Deprivation measures: The English Indices of Deprivation (2010) are provided 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).[54] 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 A11.

[54] 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 some sensitivity tests.

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.[55] 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 and we will discuss sensitivity analyses

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

in a separate subsection.

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 some big streets are not found in the 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.

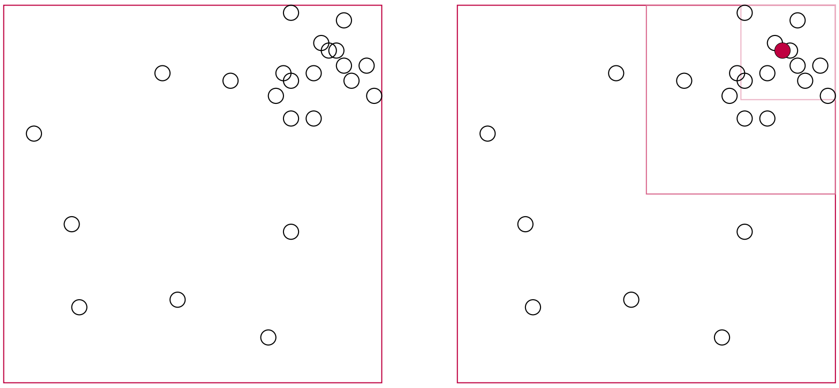
Recognizing clusters and inference We infer the geolocation of all households of the same census cluster f from the geolocation of a subsample of households (with potential measurement error). We start with the sample of well-matched households and apply the following algorithm to detect geographic clusters.

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 A15 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 but also smoothes georeferences among already-matched units.

Figure A15. Finding geographic clusters among georeferenced households in the same census cluster.



Sensitivity analysis The previous methodology relies on two approximations.

First, census clusters are assumed to reflect underlying geographic identifiers. However, there is a tension when aggregating entries together into a census cluster. On the one hand, having more households per census cluster raises the probability to detect the geographic location. On the other hand, there exist breaks within a book, and the first households may be interviewed in a neighborhood while the last households may correspond to a new interviewer and a new neighborhood. In order to alleviate this issue, we repeat our algorithm by generating different census clusters (grouping 1, 2, 3, 4, 5 pages together, drawing new breaks) and 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 a 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 right quarters, and when this number is lower than 1/2, we generate a dummy indicating that the solution to the algorithm may be subject to noise.

E Additional evidence on the nature of neighborhood effects

The analysis developed in Section 6 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 stylized facts about formerly polluted neighborhoods in 2011 which are only briefly referred to in Section 6. 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 consumptive amenities and its relationship with past atmospheric pollution.

We first rely on two specifications to better understand the correlation between the within-city distribution of consumptive amenities, past atmospheric pollution and neighborhood composition.

In a first specification (see Appendix Table A8), we report the estimates for specification (S1) where we replace our benchmark indicator of neighborhood composition by (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, however, 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 the domain *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.

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.[56]

In a second specification (see Appendix Table A10), we run specification (S1) with our benchmark indicator of neighborhood composition in 2011 and we sequentially control for public amenities (Panel A), housing supply (Panel B) and an extended set of schooling and crime indicators (Panel C). This approach complements the previous one, and implicitly provides a decomposition of the correlation between neighborhood composition and past pollution. Panel A shows that controlling for public amenities (in 1881 or 2011) does not affect the gradient between neighborhood composition and past pollution in a significant manner as expected from the results presented in Table A8. By contrast, half of the correlation is absorbed by social housing and an additional quarter disappears once we control for housing supply (flats, villas, square meters, number of bedrooms etc.). Interestingly, controlling for building age or limiting the sample to areas with a majority of houses constructed after 1940 does not affect our benchmark relationship. Finally, we control for (i) schooling supply, (ii) school composition, (iii) the presence of police forces and (iv) crime incidence in Panel C. While schooling supply and the presence of police forces do not alter the benchmark relationship, adding indicators of school composition and crime incidence captures between 20 and 40% of the initial relationship.

These descriptive statistics indicate that the persistence in segregation does not relate to rigid consumptive amenities (e.g., provision of public services, Victorian housing stock or old private schools constructed during industrialization) which would anchor neighborhoods in a certain equilibrium for decades. Instead, some consumptive amenities that correlate with past pollution could rapidly change with the composition of residents (e.g., school composition, crime incidence). While our data do not allow us to estimate the role of the different neighborhood effects, such investigation would be needed to derive policy implications that rely on targeting the provision of these amenities.

[56]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.