East Side Story: Historical Pollution and Persistent Neighborhood Sorting*

Stephan Heblich Alex Trew Yanos Zylberberg

June 18, 2017

Abstract

Why are the East sides of former industrial cities more deprived? We relate this to the historically unequal distribution of air pollutants across neighborhoods. We geolocate nearly 5,000 industrial chimneys in 70 English cities in 1880 and use an atmospheric dispersion model to recreate the spatial distribution of pollution. Individual-level census data show that pollution induced neighborhood sorting over the nineteenth century: it explains up to 15% of within-city deprivation in 1881. These equilibria persist to this day even though the initial pollution sources have waned. A quantitative model shows the role of non-linearities and tipping-like dynamics in such persistence.

Keywords: Neighborhood Sorting, Historical Pollution, Deprivation, Persistence, Environmental Disamenity.

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; yanos.zylberberg@bristol.ac.uk. We are grateful to Philipp Ager, Nate Baum-Snow, Kristian Behrens, Davide Cantoni, Peter Egger, Klaus Desmet, Christian Dippel, Gilles Duranton, Oliver Falck, Walker Hanlon, 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, USD, Warwick and participants at the 2016 IEB Workshop on Urban Economics, the 2016 EUEA, the 2016 UEA, and the 2017 ASSA meeting for helpful comments. We are grateful to Nicholas Cheras, Qingli Fang, Joanna Kalemba, Matthew Litherland, Filip Nemecek, Ondrej Ptacek for excellent research assistance. This work was part-funded by the Economic and Social Research Council (ESRC) through the Applied Quantitative Methods Network: Phase II, grant number ES/K006460/1. The usual disclaimer applies.

1 Introduction

> *"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."*

- Stephen Mosley (2013), *The Chimney of the World*

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*). However, there is surprisingly little analysis of the reasons behind this pattern. In this paper, we show that this East-West gradient is the most visible remnant of a more general process induced by the atmospheric pollution which affected cities during the Industrial Revolution. In a first step, we focus on the nineteenth century and document the relationship between the distribution of clean air within English cities and neighborhood sorting. We find that concentrations of pollution from historical factories account for 15% of the variation in neighborhood composition in 1880. In a second step, we turn to the post-pollution period and analyze the dynamics of neighborhood segregation. We find strong non-linearities in the dynamics of persistence. In cities with high dispersion in environmental (dis)amenities, tipping forces anchored formerly polluted neighborhoods. The effects of the now absent pollution are still felt to the modern day.

These findings are relevant for contemporary policy issues. First, the long-run impact of environmental disamenities on the spatial organization of cities holds important implications for the design of environmental and urban policies for economies, such as China, in the process of structural transformation. Second, many developed economies consider costly urban policies in order to open up deprived areas to new housing opportunities and business investment. We identify a tipping force in the dynamics of segregation which can inform policy makers seeking to reduce spatial inequalities.

While the impact of pollution on the environment and on the spatial organization of cities is echoed in policy debates, there is almost no work incorporating the long-run responses of economic agents. Providing such evidence is indeed challenging.

There exist no adequate data on historical pollution and data on residential sorting are equally scarce. Our empirical analysis combines detailed pollution information from 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. First, we develop an algorithm to geo-locate industrial chimneys from historical Ordnance Survey (OS) maps of the 70 largest metropolitan areas in England over the period 1880-1900.[1] Second, we use the world leading modelling system for atmospheric emissions (ADMS 5)[2] 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 novel algorithm to overcome a shortcoming of historic censuses: as data are nested in relatively large spatial units, such as ancient parishes in England, they are of limited use for within-city analysis of neighborhoods. Our algorithm geo-locates entries of the 1881 English census and matches them to low-level administrative units (for our purpose, the 2001 Lower Super Output Areas, LSOA). Across the 70 metropolitan areas, we observe 4,500 LSOAs versus 500 parishes, and cities like Bristol, Liverpool or Manchester are covered by about 90 LSOAs instead of 10 parishes.[3]

There is 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 18 percentage points in the share of low-skilled workers.

The variation in pollution exposure results from a complex interaction between the spatial distribution 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 locate a chimney, and compare upwind and downwind neighborhoods with similar topography and at the same distance from the chimney. In order to isolate such variation in the data, we proceed in three steps. In a first step, we show that the correlation is robust to the addition of a large set of controls capturing the separate effects of the distribution of pollution sources (e.g., proximity to factories, distance to amenities), topography (e.g., elevation), and wind patterns (latitude, longitude, and fixed effects at the parish or Medium Super Output Areas levels).

In a second step, we replace the actual dispersion process by counterfactual

[1]We also consider domestic chimneys but their contribution to overall pollution is small.

[2]Atmospheric Dispersion Modelling System (ADMS) models have been developed to make use of the most up-to-date understanding of the behavior of the lower levels of the atmosphere.

[3]The median LSOA in our sample covers an area of 0.3 square kilometers.

processes, generate counterfactual pollution patterns and run horse races with the actual pollution pattern. There remains one important concern: chimneys may have been selectively located upwind of poor areas.[4]

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 around 1800 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, we exploit the difference between upwind and downwind neighborhoods at the same distance from potential factories located along waterways. Our second counterfactual 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 recent years and analyze the dynamics of persistence between 1971 (just after the second Clean Air Act of 1968) and 2011.[5] Over the course of this period, pollution from coal burning abruptly decreased and yet, 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 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

[4]To reduce concerns about reverse causality or omitted variable biases, we first discard the existence of unobserved pre-existing (dis)amenities in polluted neighborhoods by looking at neighborhood composition in 1817, and running difference-in-difference specifications between 1817 and 1881.

[5]The first Clean Air Act was enacted in 1956 as a reaction to the Great Smog of 1952 in London. However, as apparent from Appendix Figure A12, the second Clean Air Act in 1968 caused a much more abrupt drop in coal consumption.

sorting which extends Lee and Lin (2013). The location choices of high- and low-income individuals depend on consumptive amenities with some amenities being tied to the neighborhood composition (e.g., through preferences, public goods accumulation or other man-made amenities). We estimate the model to match the dynamics over the period 1881-1971, and 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.[6]

Our paper makes important contributions to different strands of the literature. First, we contribute to the literature on neighborhood sorting. We show that a large but temporary environmental disamenity may modify the spatial organization of cities in the long run. To the best of our knowledge, we are the first to present evidence for this pollution-driven residential sorting in cities before, during and after industrialization (Kuminoff et al., 2013, review the existing sorting literature). Closely related papers that look at pollution-induced sorting today include Banzhaf and Walsh (2008) and Chay and Greenstone (2005). Another related paper is Redding and Sturm (2016), who use Second World War destruction in London to identify patterns of spatial sorting across neighborhoods. 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.

Second, we contribute to the literature on the dynamics of segregation and tipping points (Schelling, 1971; Anas, 1980; Card et al., 2008; Logan and Parman, 2016). Even after the sharp decrease of industrial pollution in English cities, formerly polluted neighborhoods remain the poor parts of town.[7] 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

[6]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.

[7]The phenomenon of segregation relates closely to that of gentrification which concerns the rise of historic centers in the United States (Brueckner and Rosenthal, 2009; Brueckner and Helsley, 2011; Guerrieri et al., 2013). We capture some signs of revitalization in formerly-polluted English cities but we only observe reversion to the mean for neighborhoods with initially moderate levels of pollution.

United States focusing on ethnic considerations). Moreover, we exploit a temporary disamenity to explain the initial spatial distribution of residents instead of permanent differences across neighborhoods (as present in Lee and Lin, 2013, for instance). Two recent papers find similar patterns of path dependence for neighborhoods in New York City (Villarreal, 2014) and Los Angeles County (Brooks and Lutz, 2016).

Third, our approach to modelling residential sorting builds upon Brueckner et al. (1999) and Lee and Lin (2013). Lee and Lin (2013) develop a dynamic model of household neighborhood choice to assess the role of natural amenities in sustaining the spatial distribution of income over the period 1890–2010 in the United States. Instead of natural amenities that anchor sorting, we look at pollution as an impermanent natural amenity whose effect is long-lasting due to the presence of endogenous neighborhood effects. While we use Durlauf (2004) and Rosenthal and Ross (2015) to inform our structural functional forms for neighborhood effects, our data do not allow us to disentangle underlying mechanisms.

Fourth, our paper relates to research on the spatial distribution of income and population between cities (Bleakley and Lin, 2012; Redding et al., 2011). A paper that is particularly relevant is Hanlon (2016), who argues, in the same context, that coal-based pollution was a significant disamenity with a strong negative impact on city size. We complement this research by adding a within-city perspective that shows substantial effects operating through residential sorting.

Fifth, 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 (2016) 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 predictive power of state-of-the-art pollution models to estimate historical pollution. Our third and most important methodological contribution is to provide an algorithm that geo-locates 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 geo-references of residents from a small share of well-matched neighbors.

Finally, an important line of research examines effects of pollution exposure 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 (recent contributions include (Anderson, 2015; Deryugina et al., 2016). (Graff Zivin and Neidell, 2013), 2013 review this comprehensive body of literature).

Our paper is most closely related to historical assessments of the effect of coal use on health (Clay and Troesken, 2011; Barreca et al., 2014; Clay et al., 2016; Beach and Hanlon, 2016).

The remainder of the paper is organized as follows. In Section 2, we develop a stylized model of neighborhood sorting. Section 3 briefly provides some elements of context. We detail our main data sources and methodology in Section 4, and our empirical strategy in Section 5. We analyze the relationship between neighborhood sorting and historical pollution in Section 6. Section 7 looks at the dynamics of persistence between 1971 and 2011 and relies on a quantitative, dynamic version of the model developed in Section 2. Finally, Section 8 briefly concludes.

2 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 7.

In the model, neighborhood sorting arises out of within-city differences in consumptive amenities.[8] Neighborhoods are made up of an interval of locations, and the amenity level at a location is part air quality (constant within a neighborhood) and part location-specific. As in Lee and Lin (2013), there is a complementarity between consumption and amenities. Willingness to pay rent in high amenity locations is relatively higher for the high income types: in equilibrium, all high income workers are housed in the best amenity locations, all low income workers are in the other locations, and a difference in air quality 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 $\mu(\Omega(j)) = 1$, and 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 7 will be dynamic, we assume here a

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

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 consumption and amenities are complementary, 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).[9] We assume that $x(\ell,j)$ is, in both 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 $\bar{V}^l = 0$ and the rent charged to a low-skilled worker is $R^l(j,\ell) = \theta^l$ 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

[9] 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.

$I^l(j,\ell) = 1$ if location ℓ in neighborhood j is rented to a low-skilled worker, the land-worker clearing condition is,

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

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 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 A. □

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 A. □

The impact of pollution is depicted in Appendix Figure A2. The disamenity causes equilibrium rents paid by high-skilled workers to increase compared to the benchmark without pollution. Since the lowest 2γ 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 (and topography).

3 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), and the transition to coal as a dominant energy source occurred only after the 1840s.[10] This late energy transition is reflected in Appendix Figure A12: 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 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 Figure A12, 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. 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 the intensity of 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 is also reflected in the rapid recovery of the *Biston betularia insularia* between 1970 and 2000 (Cook, 2003).

[10] 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.

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 (see Shaw-Taylor and Wrigley, 2014). As shown in Williamson (1990) and the Appendix Figure A12, 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.[11]

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.

4 Data

This section describes the construction of our measures: 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 geo-locate the 1881 Census.

4.1 Construction of the Air Pollution measure

Our strategy to generate a geo-referenced air pollution map for 70 metropolitan areas covered by Ordnance Survey (OS) maps can be summarized as follows. In a first step, we go through each geo-referenced map tile and mark each chimney with a unique identifier. We use a recognition algorithm to locate each mark, and extract the associated identifier. In a second step, we predict atmospheric dispersion of polluting particles from each individual chimney and we isolate a chimney-specific pollution imprint. In a third step, we consider a relevant geographic unit, i.e., 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.

[11] 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.

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 large white circle, and they are 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 I). 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.[12] The red symbol X is located in the center of a chimney and is used to geo-locate the chimney. The symbol X can be 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 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* while *00006* belongs to *Britannia Saw Mills*.

We restrict our analysis to 70 metropolitan areas in England (see the online Appendix Figure A14). These cities constitute a quasi-exhaustive snapshot of industry and its associated pollution, and cover between 60% and 66% of the total population in 1801, 1881 and 2011.

Dispersion modelling Atmospheric dispersion is calculated using the *ADMS 5* dispersion modelling software.[13] This model is an augmented version of the basic Gaussian air pollutant dispersion equation known as the Gaussian-Plume model. In addition to the standard Gaussian-Plume model, *ADMS 5* includes a wide variety of options, some of which are directly useful in our context. *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. Another feature that is particularly important in our context is to account for complex terrain and the changes in surface roughness. Since in-

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

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

dustrial 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 the contemporary 10-year statistical meteorological data as provided by the Met Office for the different cities in our sample, thereby neglecting potential changes related to climate fluctuations between the 19th century and today. Second, the model requires complex terrain data and convective meteorological conditions on land. We use the current topography, such as terrain height and roughness which affects 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 high, wide and have high exit velocity. By contrast, chimneys in the Industrial Revolution were between 10 and 50 meters high, most being lower than 25 meters. Moreover, the exit velocity and temperature were also lower than today's chimneys. After discussing with *ADMS 5* modellers, we decided to use a point source (i.e., an emission from a stack) that is 25-meters high, with an exit velocity of 4 m/s and an exit temperature of 120 degrees Celsius.

For illustration purposes, we report in Appendix Figure A9 the "wind roses" which indicate wind provenance and intensity for Northern England and Southern England. Local wind stability is also important and, as apparent, wind is much less predictable in Northern England generating more disperse air pollution measures on average (over a 10-year period). In Appendix Figure A10, we show the differences in pollutant dispersion implied by topography in Halifax and Oldham.

We report one sensitivity check for illustration purposes in the online Appendix. In Appendix Figure A11, we plot the air pollution measures as a function of distance to the source under stable and unstable conditions and for three different chimney heights. 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 predominant direction). Note that our benchmark measure uses an average of these conditions over the past 10 years.

We also model pollution related to domestic emissions. To this end, we use a volume source and we consider the emitters as being uniformly distributed within the city borders (as drawn in the Ordnance Survey maps). We use the same meteorological and topographic inputs as for the industrial emissions.

[14] Topography and land cover play little role in flat terrains.

Aggregation at a given geographical unit Atmospheric dispersion models are additive. Total Air Pollution measures are computed as the sum of 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 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 cloud by this industry-specific coefficient to derive an aggregate measure of air pollution. We finally collapse our data at the level of 2001 Lower Super Output Areas to assign our pollution measure to administrative units. Figure II 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.

4.2 Geo-locating 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, and 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 the Appendix B.

The intuition behind our methodology is the following. There are two indicators of household location: A geo-located 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 and they filled in enumerator books while visiting these neighborhoods. As a result, there is a strong clustering among census entries.[15] If we locate a fraction of households, we can infer the geo-references 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.

[15] 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".

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.[16]

To implement our cluster analysis, we need to geo-locate 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 geo-located addresses. We then run a fuzzy matching procedure between the pool of census addresses and the pool of contemporary geo-located addresses within the same parish of registration. We achieve a perfect match for about 20% of the total sample, and we match 30% of the total sample with precision 0.90 (i.e. at least 90% of the original string can be found in the matched address).[17]

Clustering algorithm A precise description of the algorithm is provided in Appendix B 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 (2011 Lower Super Output Area - LSOA). 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 likely LSOA identifier.

Neighborhood composition For 1817, we use "The Occupational Structure of England and Wales, c.1817-1881" (Shaw-Taylor and Wrigley, 2014) which use baptism records over 1813–20 to reconstruct a quasi-census for 1817. The resulting 834 parishes can be consistently linked with parishes in 1881. For 1881, we use the census at that year merge with LSOA units.

For recent waves (1971–2011), we use census aggregate data at the 2001 Lower Super Output Area (LSOA) level 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.

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

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

One drawback is that 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 present in the recent censuses, and rely on a similar classification (PST system of classifying occupations; see Wrigley (2010)) constructed by The Cambridge Group for the History of Population and Social Structure for 19th century censuses (1817 and 1881).

There exist many proxies for income based on occupational structure. For instance, one could infer the synthetic LSOA income from average occupational wages 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 in our benchmark analysis on a proxy based on the raw data, i.e., the LSOA share of low-skilled workers among the working population.[18]

For 1817 and 1881, we first collapse 500 occupational categories into 10 categories. We define low-skilled workers as the Unskilled and Semi-Skilled workers. We classify Managers, Gentlemen, Rentiers, Clerks, and Manual Skilled workers as high-skilled workers. Finally, we assign Farmers to a separate category and we drop Soldiers and the Disabled 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.[19] This decomposition covers about 60% of low-skills, 30% of high-skills and 10% of farmers in 1881 (resp. 78%, 12% and 10% in 1817) in our 70 metropolitan areas.

For 1971-2011, the occupational categories are already classified into 1-digit clusters: Managers; Professionals; Associate Professionals; Administration; Manual Skilled; Care; Sales; Processing; and, Elementary. We replicate our classification in the main categories for males between 25 and 55 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 between 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 modern, urban LSOAs. We also consider alternative indicators of neighborhood composition for the recent period such as house prices or the different components of deprivation as compiled in the English Indices of Deprivation (2010).

[18]We can alternatively normalize this share by the share of low skilled workers in the city. Doing so does not affect our results.

[19]Our results are robust to (i) adding female workers, and (ii) widening the age interval (e.g., 15-65).

4.3 Descriptive statistics

The Appendix Table A1 provides summary statistics for the full sample and, for each variable, a decomposition of the variance within and between cities. The clustering process applied to neighborhoods within a buffer of 20 kilometers around the centroids of our 70 cities classifies about 12 million individuals in $4,524$ LSOAs in 1881.[20] As these LSOAs are the 2001 census units, we can associate contemporary measures for all $4,524$ observations, and we only lose 5 LSOAs when we create topographic controls.

In the first line of Appendix Table A1, we report summary statistics for our normalized pollution measure. As apparent from the last columns, a very large share of the variance in the pollution measure is within cities. Our empirical strategy hinges on such within-city variation and is mostly orthogonal to between-cities variation. The share of low-skilled decreased from 78% in 1817 to 61% in 1881 and, as before, a significant share of the variance is within cities. Finally, we also report descriptive statistics for our most important geographic and topographic controls.

To better understand the extent to which cities were polluted at the end of the nineteenth century, we provide the cumulative distribution for non-normalized pollution in our sample of LSOAs. Figure A13 shows that about 10% of our sample LSOAs display air pollution above the National Ambient Air Quality Standards (SO2 concentration above $12-15\mu g/m^3$). About 2% of our sample LSOAs – mostly in Manchester, Oldham and Liverpool – have indices of pollution above the peaks recorded in contemporary Beijing ($40\mu g/m^3$).

We also provide in Appendix Table A9 an illustration of the within-city variation in air pollution. We compare our estimates to 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.[21] 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, our estimates strongly correlate with the deposit measure (correlation of 0.92).

[20] Large cities such as Bristol, Liverpool or Manchester are covered by almost 100 LSOAs each. With fewer restrictions on the clustering process or the fuzzy matching, our sample size would slightly increase at the expense of including LSOAs with higher measurement error.

[21] 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.

5 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 a LSOA, p a parish, c a city, and t a particular census wave ($t = 1881, \ldots, 2011$), we estimate the following equation:

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

where Y_{it} is a measure of occupational structure. As P_i, the actual pollution, is a complex interaction between the distribution of pollution sources and a dispersion process, we include separate indicators for topography (e.g., elevation), geographic controls (e.g., latitude and longitude) and proximity to factories in the set of controls $\mathbf{X}_i$ (see the full list in Appendix Table A2). Y_p is the occupational structure in 1817 at the parish level and δ_c are city Fixed-Effects.

We further explore the interaction between the distribution of pollution sources and the dispersion process by considering alternative dispersion processes, e.g., synthetic pollution generated by the same pollution sources but a "static" wind. Doing so, we decompose the interaction such as to isolate variation induced by the asymmetry between neighborhoods equidistant from factories, some of them being located downwind or upwind.

Another concern with specification (S1) is that the treatment may not be exogenous because some unobserved amenities may explain both the upwind presence of industries and the occupational structure in some neighborhoods. In robustness checks, we show a balance test in 1817, and run a difference-in-difference specification identifying the sorting in response to pollution from differences between the pre-treatment period (1817) and the post-treatment one (1881).[22]

There remains a threat to identification. Factories may have been strategically placed upwind of poor neighborhoods such as to minimize political or economic costs associated with environmental disamenities in richer neighborhoods.

Controlling for non-random industry location In order to clean the variation in pollution from variations due to non-random industry location, we predict pollution imprints that draw on exogenous variation in pollution sources. In other

[22] Our unit of analysis in 1817 is the parish and a proper difference-in-difference specification therefore requires using the parish as the unit of observation throughout. Since cities like Bristol, Liverpool or Manchester are covered by about 10 parishes, we lose a lot of information.

words, we interact an exogenous variation underlying the choice of industry location with the atmospheric dispersion due to wind flows and topography.

We suggest two different ways to obtain exogenous variation in industry location based on the city structure. In a first specification, we exploit the fact that large boilers required a constant stream of water for cooling. As a result, all mills were located along rivers or canals (Maw et al., 2012). We locate synthetic chimneys uniformly along natural waterways in 1827, before the rise of coal as the main energy source, we assume uniform air pollutant emissions from these counterfactual sources and use the same atmospheric dispersion model.[23] These synthetic industry locations were not susceptible of being selectively placed upwind of poor neighborhoods because of resulting pollution. However, this variation also correlates with distance to waterways which may itself affect the attractiveness of a neighborhood. We thus control separately for this distance and only use the asymmetric atmospheric dispersion as predicted by the existing meteorological conditions and topography.

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). We assume uniform air pollutant emissions from the centroid of each industrial district and we use the atmospheric dispersion model to create the counterfactual pollution pattern. The underlying assumption is that steam engines are a reasonable proxy for the historical center of industrial districts whose location was chosen before the rise of large-scale coal pollution. As in the previous case, we separately control for distance to the nearest historical district, and only rely on an upwind/downwind pattern in pollution around each industrial district.

We then use following two-stage specification to isolate the residual air pollution predicted by the synthetic air pollution imprint to estimate its effect on sorting:

$$\begin{cases} P_i = b_0 + b_1 PP_i + \mathbf{c}\mathbf{X}_i + d_c + fY_p + e_{ict} \\ Y_{it} = \beta_0 + \beta_1 \widehat{P}_i + \gamma \mathbf{X}_i + \delta_c + \nu Y_p + \varepsilon_{ict} \end{cases} \tag{S2}$$

where Y_{it} is a measure of occupational structure, PP_i is one of the two synthetic treatments predicted by chimneys located along the 1827 natural waterways or steam engine locations, P_i is the pollution as predicted with actual chimney locations, $\mathbf{X}_i$

[23] The Appendix Figure A4 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 counterfactual chimney locations and panel (c) the resulting air pollution. Finally, panel (d) shows the actual pollution.

are controls including distance to waterways, Y_p is the occupational structure in 1817 at the parish level and $\{\delta_c, d_c\}_c$ are city Fixed-Effects.

6 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 at least 15% of the social composition across neighborhoods of the same city. While we control for important neighborhood characteristics in our benchmark specification (longitude and latitude, distance to main amenities including waterways 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 control for counterfactual atmospheric pollution clouds to condition our analysis on neighborhoods being on the same ring around a factory. Third, we run our two-stage specification to isolate exogenous variations in chimney locations.

Benchmark results In Table I, 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, and the correlation is both statistically and economically significant. The coefficient is precisely estimated and the 95%-confidence interval is $[.028, .055]$. One additional standard deviation in air pollution increases the prevalence of low-skilled workers by 4.2 percentage points, which is about 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 18 percentage points in the share of low-skilled workers.

Controlling for a large sets of covariates does not affect our baseline estimates. In the second column of Table I, we add city fixed-effects to control for variation in atmospheric pollution and occupations between cities.[24] In the third column, we add 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 the separate elements entering in the pollution dispersion process. In the fourth column, we condition our estimates on topography (mean/min/max

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

elevation and distance to waterways in 1827). In the fifth column, we control for proximity to industries with the number of chimneys in the area, distance to the city hall, area and the share of the LSOA within the 1880 city borders.[25] In the sixth column, we add latitude and longitude of the LSOA centroids to control for wind patterns and potential Western or Southern preferences in locations. As apparent from Table I, our estimates slightly decrease (from 3.8 to 3.3 percentage points in low-skilled workers per standard deviation in air pollution) but remain large and significant at the 1% level.[26]

Figure III 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 4 of Table I. 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.

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

Robustness checks We run a series of robustness checks to control for pre-pollution neighborhood composition and distance to factories, and to ensure that our estimates are not driven by non-random and strategic location of industries within cities.

First, we account for fixed LSOA characteristics in Table II and Appendix Table A3. Panel A of Table II shows a "balance test." We estimate 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 five specifications, the coefficient is not different from 0 statistically and economically. This placebo check is reassuring since it suggests that potentially unobserved, pre-existing neighborhood characteristics are not driving our results. 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 II.

In Appendix Table A3, we run difference-in-difference specifications either at the

[25] As stated in Section 4, the metropolitan areas (and thus the sample of LSOAs) are defined by a buffer of 20 kilometers around the centroids of our cities. In unreported robustness checks, we verify that the results are left unchanged if we limit the sample to a 10 kilometers buffer, a 5 kilometers buffer or all LSOAs intersecting with the 1880 city borders (which may be endogenous and affected by pollution through the relative returns to agriculture).

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

LSOA-level (attributing the average parish-level share of low-skilled workers in 1817 to LSOAs) or at the parish-level. As shown in Panel A, the difference-in-difference estimates at the LSOA-level are very similar to the simple difference estimates of Table I. The estimates for the difference-in-difference specification at the parish-level (Panel B) are slightly larger, but less precisely estimated. This approach, along with the balancing test, reduce concerns about biasing effects from fixed unobserved LSOA 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 third set of robustness checks.

Third, we generate three counterfactual pollution imprints from actual industry locations but alternative air pollution profiles.[27] In order 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 the wind profile *rotated* in steps of 30 degrees around each source. Figure IV illustrates the results for the years 1817 and 1881 where we center the figure around the actual pollution pattern at 0 degrees. In 1817, *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, *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 the wind patterns, the estimated relationship looses significance and turns negative. The fuzzy relationship between 0 and 30 degrees may be due to measurement error.[28]

Next, we control for the density of chimney around a neighborhood. Note that 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 the synthetic atmospheric pollution from existing chimneys under a static wind profile, symmetric in all directions (*Static* pollution). This measure is additive in the number of chimneys and thus also captures the proximity to a dense cluster of factories. As shown in Table III–column 1, this counterfactual atmospheric pollution measure does not affect our estimates.[29] In column 2, we include a placebo industry pollution

[27] 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 the coal consumption of the industry gives the same results. Our estimates are quantitatively and qualitatively similar with these more conservative chimney weights.

[28] 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.

[29] As the measure of *static* pollution is positively correlated with the share of low-skilled workers

pattern that assumes low pollution values for chimney in high-pollution industries and high pollution values for chimney in low-pollution industries. Including the measure does not affect our estimates and we observe a negative and insignificant coefficient on the placebo pollution measure. In column 3, we control for domestic pollution as predicted by the location of private residential buildings across the city, which has a small predictive power and again does not affect our main estimate. Finally, we add SO2 concentration in 2011 as measured by the Department for Environment, Food and Rural Affairs (DEFRA). Due to deindustrialization and new sources of emissions, the within-city correlation between past and contemporary pollution is low, and our main estimate is left unchanged (column 4).

Fourth, we present in Table IV the results of our 2-stage 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 similar estimates: the 2-stage estimates tend to be larger than the OLS estimates whether we control for geographic coordinates (columns 2 and 4) or not (columns 1 and 3). One additional standard deviation in air pollution increases the prevalence of low-skilled workers by about 6 percentage points in the 2-stage strategy with 1827 natural waterways, and 9 percentage points in the 2-stage strategy 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 for the downward biased results could be measurement error or possibly the difference between the local average treatment (relying on small variations around the mean within-city pollution) and the average treatment. Following this exercise, we focus on the OLS coefficients in what follows noting that they might be a conservative lower bound.

Finally, we present in Appendix Table A4 a sensitivity analysis for three elements of our baseline specification: the choice of fixed effects, clusters, and sample selection. In Panel A, we report the results of our baseline specification (Table I—column 4) with parish-fixed effects (column 1) instead of city-fixed effects. We further expand our set of fixed effects in column 2 to electoral wards (1270 in our sample) and in column 3 to Medium Lower Super Output Areas (1600 in our sample). While the correlation slightly decreases with this largest set of fixed effects, the estimates remain non-negligible even when identification comes from a within-MSOA comparison.[30] In Panel B, we report standard errors clustered at three different levels,

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.

[30]There are 3 Lower Super Output Areas on average within each MSOA in our sample.

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 conclusion, our findings do not disappear once accounting for spatial auto-correlation in atmospheric pollution. In Panel C, we estimate the baseline specification on alternative samples. In column 1, we exclude the London region. We exclude the North-West including Manchester and Liverpool in column 2, and the North-East in column 3. In all cases, the estimates slightly fluctuate around the baseline but these variations are not quantitatively relevant.

7 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 pollution. We then analyze the dynamics of persistence between 1971 and 2011.

7.1 Historical pollution and contemporary neighborhood segregation

In a first step, we expand on our previous analysis of neighborhood sorting in 1881 to recent waves in 1971, 1981, 1991, 2001 and 2011. Table V reports the slopes between the shares of low-skilled workers for the different waves and historical pollution. One additional standard deviation in historical air pollution increases the prevalence of low-skilled workers by 3 to 3.5 percentage points, and the standardized effects range between .18 and .22 without a clear time pattern between 1971 and 2011. A differential in pollution equivalent to the one between the first and last deciles in Manchester is still associated with a differential of 14 percentage points in the share of low-skilled workers, thereby explaining the persistence of the social gradient between West and East often evoked in popular culture.

We present in the Appendix two robustness checks with (i) alternative measures of deprivation and (ii) house prices. In Panel A of Appendix Table A5, we use deprivation indices (income, employment, education, health, barriers to housing and crime).[31] The estimates are very large and significant: one additional standard deviation in air pollution increases the deprivation scores by .20 to .30 standard deviations for the income, employment and education sub-indices. The only index that is not correlated with past pollution is the one capturing physical and financial accessibility of housing and local services. In Appendix Table A6, we use transac-

[31]See a description of data sources in Appendix C.

tions in England and Wales between 2009 and 2013 as recorded by Nationwide (a major mortgage provider) or Land Registry (an administrative registry of all house transactions) and create the (logarithm of the) average transaction prices and number of transactions between 2000 and 2011. The first (resp. last) columns of Table A6 report estimates for house prices (resp. transactions) and we present two specifications, one without house controls and one controlling for house characteristics. We find that one additional standard deviation in past pollution is associated with an unconditional price drop of about 10% (19% of a standard deviation). Controlling for house types provides a conditional price malus of about 5 to 8% (9 to 12% of a standard deviation). Looking at the number of transactions, we find that an additional standard deviation in past pollution is associated with a 8% (14% of a standard deviation) decrease in the number of transactions and 15% (28% of a standard deviation) when controlling for house characteristics. We illustrate the relationship between house prices and pollution (as estimated in Table A6–column 1) in the Appendix Figure A5.

The persistence of the relationship between historical pollution and neighborhood composition cannot be mechanically attributed to the collapse of industries in the former *cottonopolis*. Indeed, our estimates are identified within cities, and are robust to controlling for distance to factories, a larger set of fixed effects or the 2-stage specification (unreported specifications).[32]

In order to visualize non-linearities in the persistence of neighborhood sorting, we display in Figure V the relation between the shares of low-skilled workers in 1881 (dashed line), 1971, 1981, 1991, 2001 and 2011 (plain lines) and the temporary disamenity. As apparent, we observe a reversion to the mean for intermediate values of within-city pollution. By contrast, returns to the mean for extreme values of pollution, i.e., one standard deviation above or below average within-city pollution, are much less pronounced between 1971 and 2011. The gap with average neighborhood composition remains quite constant at the extremes. This pattern would be consistent with the existence of tipping points leading to a higher persistence in neighborhoods with the most extreme pollution exposure.

The remainder of this section will build on a quantitative model to characterize the non-linear persistence between past disamenity and neighborhood composition, and we use the model to better understand the dynamics of segregation across cities.

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

7.2 A quantitative and dynamic model of sorting

In Section 2, we laid the foundations of the quantitative model in a static framework of neighborhood sorting. We extend the model to a dynamic framework where the persistence of sorting is rationalized by an endogenous amenity.[33]

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]

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 $1-\delta$ (to be estimated), 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, $\bar{\theta}(j,t)$[35]. 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. We estimate the depreciation parameter, the characteristics of the endogenous perturbations and, if it exists, the threshold. 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 one neighborhood. The endogenous amenity, $d(j,t)$, is defined for $t > 1$ as,

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

with an initial endogenous amenity that is constant across all neighborhoods, $d(j,1) = d$. The amenity evolves according to δ, $e(j,t)$ and $b(j,t)$. The continuous perturba-

[33]In the model, workers can move freely across neighborhoods in any period, such that the persistence of sorting does not derive from frictional housing markets.

[34]Durlauf (2004) and Rosenthal and Ross (2015) are excellent overviews of the range of neighborhood effects. 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).

[35]In particular, $\bar{\theta}(j,t) = \frac{S^l(j,t)\theta^l + (1-S^l(j,t))\theta^h}{\gamma\theta^l + (1-\gamma)\theta^h}$.

tion improves amenities when a neighborhood has higher-than-average income,

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

if $\bar{\theta}(j,t-1) > 1$ and 0 otherwise.[36] The tail effect captures a discontinuity in neighborhood income, and detracts from the local amenity according to,

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

if $S^l(j,t-1) \geq \bar{S}$ and 0 otherwise.[37] 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 will persist. If neither channel operates, there is no sorting once pollution ceases.

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

Proof. See Appendix A. □

We identify the model in the data using the within-city residuals of low-skill share between 1881 and 1971 as well as the within-city residuals of atmospheric pollution for the 4,519 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 $d(j,t)$,[38]

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

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

[36] 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.

[37] 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 $\tilde{\phi}_1^b \equiv -\phi_1^b \left[\frac{(\theta^l - \theta^h)}{\gamma\theta^l + (1-\gamma)\theta^h}\right]^{\phi_2^b}$.

[38] 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 $S^l(j,t_c) = \gamma + \text{sign}\{p(j)\}d(j,t_c)/2$.

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 VI, 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.[39] 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 simulate the model using our pollution estimates for 1881 over a grid of the six model parameters (ϕ_1^e, ϕ_2^e, ϕ_1^b, ϕ_2^b, $\bar{S}$ and δ) and 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] In Table VII, 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

[39]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.

[40]We start with a coarse grid over the whole range and then increase accuracy in the grid around the initial estimates (see Note to Table VII).

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 4,519 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 1971 and 2011,

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

where $t_1 = 1971$ and $t_2 = 2011$.

The first two columns of Table VIII report results on these statistics in the model and the data. The model performs well in matching the target spread of low-skill workers 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 A8.

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

Liberalization of social housing Before 1979, social housing was distributed relatively uniformly across neighborhoods (see the online Appendix Figure A6).[41]

[41] The United Kingdom initiated a program of social housing with the Housing of the Working Classes Acts (circumscribed to London in 1890 and extended to all councils in 1900). Council housing was the main supply of housing services for the working class, and it was typically managed

In 1979, Thatcher offered social housing tenants the 'Right to Buy' their property, which endogenizes the distribution of social housing.

We model social housing as a disamenity[42] and as being occupied by low-skill workers. 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 workers. 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 households (a process which is, in the data, completed by 1991).

The third column of Table VIII reports the model output with social housing liberalization ('SH-L' in the Table) against the baseline model and the data. The liberalization removes support for low-skill workers in otherwise desirable neighborhoods; some low-skilled workers choose to sell the now-valuable housing to high-skilled workers. 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.

The analysis of social housing in the data strongly supports this interpretation. We use the Census in 1971, 1981, 1991, 2001 and 2011 and extract a LSOA-specific share of households living in council housing. In order to study the realignment of social housing with deprivation, we analyze the dynamics of social housing in formerly polluted neighborhoods (see Table A7). While social housing was weakly correlated with past pollution in 1971, it became increasingly present in formerly-polluted areas. We find that social housing already appears more in formerly polluted areas in 1981, two years after the deregulation, and it seems to reach a steady-state after 1991. In parallel, the home-ownership rate experiences a relative decrease in the areas that were formerly affected by coal pollution.[43]

While the original intent of Thatcher's policy was to reduce inequality by pro-

by local councils. About 30% of our sample of urban households were living in council houses in 1971.

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

[43] We also report the correlation between past pollution and the share of immigrants in Table A7. We find that the share of immigrants steadily increases in formerly-polluted areas with a sharp acceleration between the last two waves. In 1971, an additional standard deviation in past pollution increases the share of immigrants by about 1 percentage point against 3.5 percentage points in 2011 (about .25 of a standard deviation in both cases).

viding 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. Our estimates suggest that about 20% of the remaining gradient between polluted and spared neighborhoods can be attributed to this reform.

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 30% to 40% or 45%. As can be seen in columns 2 and 3 of Table IX, even a substantial investment in social housing would have been ineffective in reducing significantly 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 back 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.[44] As can be seen in columns 4 and 5 of Table IX, higher (lower) initial pollution increases (decreases) the spread of low-skill workers 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 differ widely in exposure to pollution. We define $share_{polluted}$ as the city-wide share of areas with pollution above the sample average and divide cities in two groups of equal size: those with a high share of polluted areas (78%, and the share of low-skilled workers in 1881 is 65%) and cities with a low share of polluted areas (25%, and the share of low-skilled workers in 1881 is

[44] This increases the gap in disamenities between the East and West neighborhoods by ±25%.

60%).

We report in Table X the empirical measures of persistence across time for the two sets of cities. As apparent in Panel A, there is reversion to the mean in cities with a low share of polluted areas. The standardized effects of 24% in 1971 drops to 16% in 1981 and 10% from 1991 onwards. By contrast, cities with a high share of polluted areas (Panel B) do not experience any reversion to the mean. The standardized effects are around 25% in 1971, and then range between 23 and 29% in the following waves. The model estimates for the correlation with pollution in 2011 are .0340 (versus .0409 in the data) for cities with a high share of polluted areas, and .0162 (versus .0175 in the data) for cities with a low share of polluted areas. In other words, the model may slightly under-estimate the persistence in very affected cities, but performs well in capturing variations across cities. The Appendix Figure A7 provides a graphical illustration of the dynamics of persistence in the two sets of cities, and the distribution of pollution within these cities.

A shortcoming of our 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, Appendix C provides detailed descriptive statistics about the distribution of a wide range of neighborhood characteristics within cities in 2011. We find that housing characteristics, schooling outcomes and crime incidence differ markedly in formerly polluted neighborhoods. However, these patterns cannot be attributed to the mere anchoring of durable amenities established before the Clean Air acts such as the provision of some public amenities (parks, public administration or transport), the presence of private schools, or the heritage of Victorian housing stock.

8 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. Historical pollution explains 15-20% of the spatial distribution of deprivation today and our results are robust to a number of alternative samples and specifications.

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.

Our findings hold at least two important implications. First, we show that the success of urban policies to revitalize deprived areas crucially depends on their position relative to the tipping point. 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 is a significant long-run consequence of an uneven pollution exposure across space: pollution may induce large spatial inequalities that survive de-industrialization.

References

Anas, Alex, "A model of residential change and neighborhood tipping," *Journal of Urban Economics*, 1980, *7* (3), 358–370.

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.

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.

Bleakley, Hoyt and Jeffrey Lin, "Portage and Path Dependence," *The Quarterly Journal of Economics*, 2012.

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

Brueckner, Jan K and Robert W Helsley, "Sprawl and blight," *Journal of Urban Economics*, 2011, *69* (2), 205–213.

_ **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.

Brueckner, Jan K., Jacques-François Thisse, and Yves Zenou, "Why is central Paris rich and downtown Detroit poor?: An amenity-based theory," *European Economic Review*, 1999, *43* (1), 91 – 107.

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.

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 "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.

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.

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

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

Guerrieri, V., D. Hartley, and E. Hurst, "Endogenous Gentrification and Housing Price Dynamics," Working Paper 2013.

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*, 2016, *forthcoming.*

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," *Federal Reserve Bank of Philadelphia Working Paper 13-48*, 2013.

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.

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.

_ , Sturm Daniel M., and Nikolaus Wolf, "History and Industry Location: Evidence from German Airports," *The Review of Economics and Statistics*, 2011, *93* (3), 814–831.

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.

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

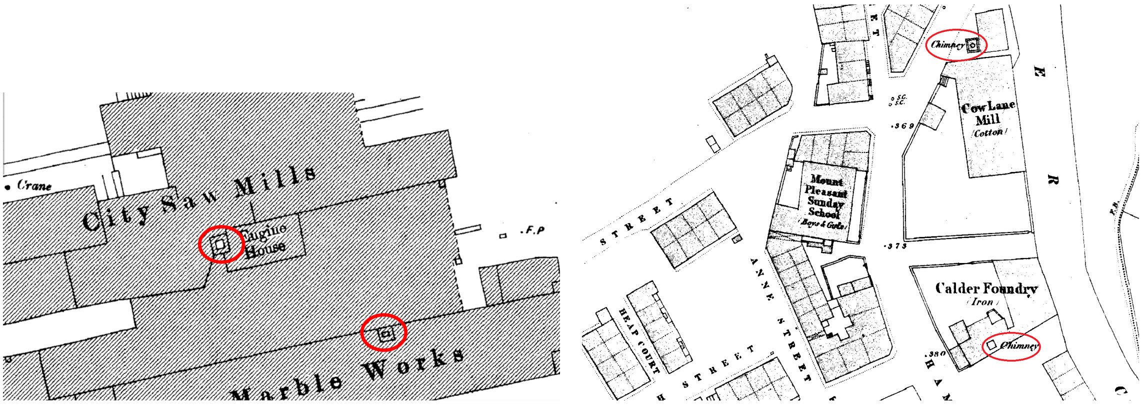
Williamson, Jeffrey G., "Earnings Inequality in Nineteenth-Century Britain," *The Journal of Economic History*, 1980, *40* (3), 457–475.

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.

A Figures and tables

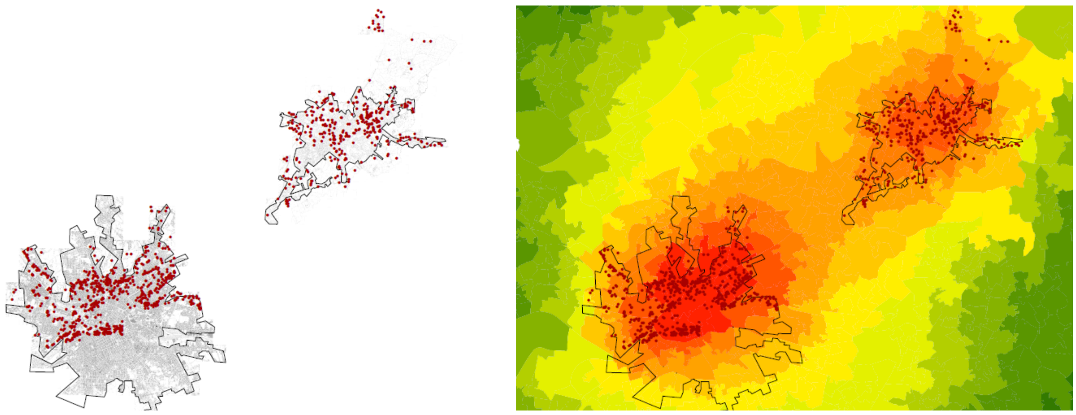
Figure I. Ordnance Survey maps – chimney symbols.



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

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

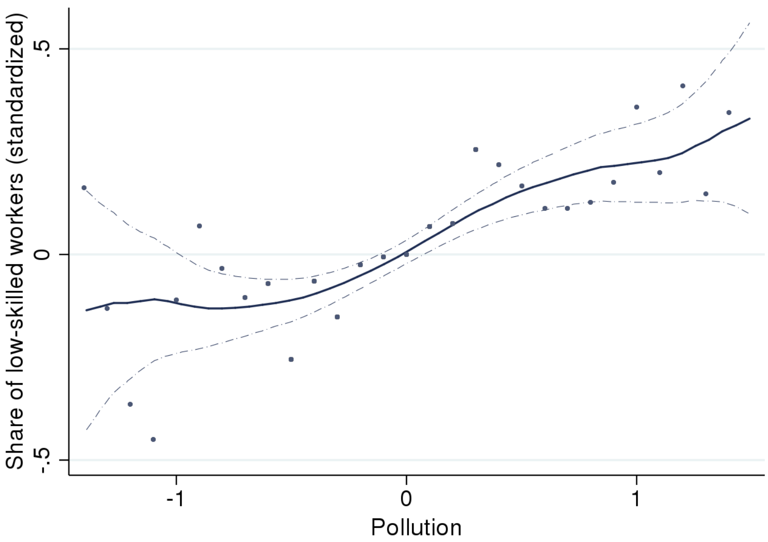
Figure II. Aggregating pollution sources (Manchester).



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

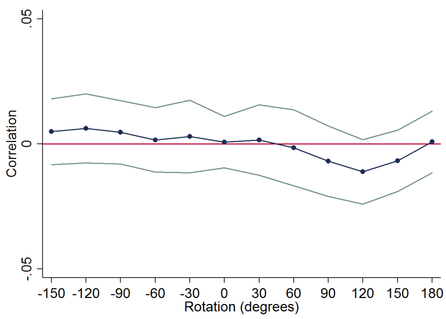
Sources: Authors' calculations using Ordnance Survey Maps - 25 inch to the mile, 1842-1952 and the ADMS 5 Air Pollution Model. Chimneys are indicated with red dots.

Figure III. Share of low-skilled workers and pollution across neighborhoods in 1881.

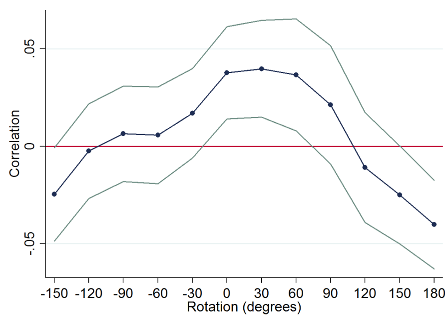


Notes: This Figure represents the relationship between the (standardized) shares of low-skilled workers in 1881 and our (standardized) measure of past pollution. We consider the residuals of all measures once cleaned by city Fixed-Effects, geographic and topographic controls. We create 40 bins of neighborhoods with similar past pollution and represent the average shares of low-skilled workers within bins. The lines are locally weighted regressions on all observations.

Figure IV. Rotating wind patterns in 1817 and 1881.



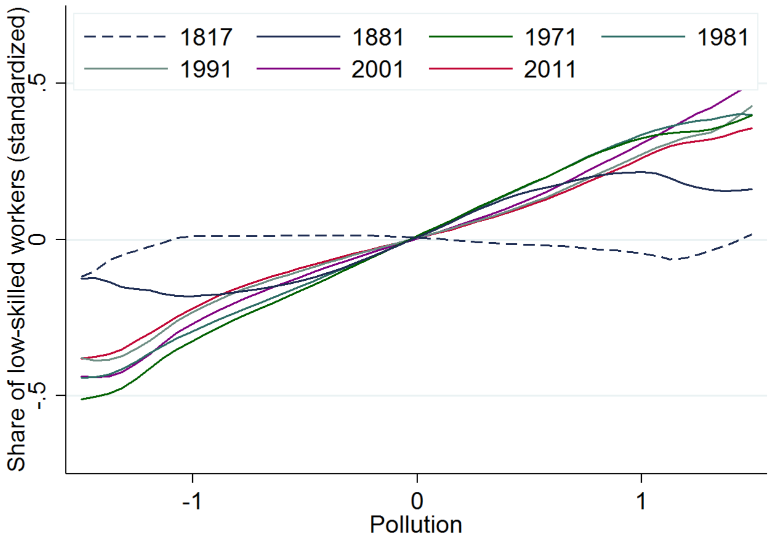
(a) 1817.



(b) 1881.

Notes: These Figures represent the correlations between the (standardized) shares of low-skilled workers and the measures of past pollution rotated in steps of 30 degrees for the years 1817 and 1881. All regressions include the controls reported in Table I, Column 6, with standard errors being clustered at the parish-level.

Figure V. Share of low-skilled workers (y-axis) and pollution (x-axis) across neighborhoods 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, geographic and topographic controls.

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

Share of low-skilled	(1)	(2)	(3)	(4)	(5)	(6)
Pollution	.0417	.0421	.0379	.0350	.0348	.0326
	(.0070)	(.0068)	(.0065)	(.0063)	(.0073)	(.0076)
	[.1686]	[.1700]	[.1532]	[.1415]	[.1405]	[.1318]
Observations	4,524	4,524	4,524	4,519	4,519	4,519
Fixed effects (city)	No	Yes	Yes	Yes	Yes	Yes
Controls (1817)	No	No	Yes	Yes	Yes	Yes
Controls (topography)	No	No	No	Yes	Yes	Yes
Controls (industry)	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 topographic controls include the average, maximum and minimum elevations for the LSOA and the distance to waterways as of 1827. The set of 1817 controls include the parish-level shares of farmers, managers and blue-collar workers. The set of industry controls include number of industrial chimneys, distance to the city hall, share of LSOA within the city borders in 1880 and the LSOA area.

Table II. Pollution and shares of low-skilled workers or wealth measures before pollution – balance tests.

Panel A: Share of low-skilled in 1817	(1)	(2)	(3)	(4)	(5)
Pollution	.0000	-.0048	.0112	.0023	.0029
	(.0125)	(.0196)	(.0263)	(.0273)	(.0269)
	[.0004]	[-.0427]	[.0988]	[.0203]	[.0260]
Observations	480	480	480	480	480
Fixed effects (city)	No	Yes	Yes	Yes	Yes
Controls (geography)	No	No	Yes	Yes	Yes
Controls (topography)	No	No	No	Yes	Yes
Controls (lat./lon.)	No	No	No	No	Yes
Panel B: Wealth in 1815	(1)	(2)	(3)	(4)	(5)
Pollution	.3795	.2322	.1428	-.0449	-.0637
	(.1472)	(.1487)	(.1563)	(.1559)	(.1654)
	[.3000]	[.1838]	[.1131]	[-.0355]	[-.0504]
Observations	450	450	450	450	450
Fixed effects (city)	No	Yes	Yes	Yes	Yes
Controls (geography)	No	No	Yes	Yes	Yes
Controls (topography)	No	No	No	Yes	Yes
Controls (lat./lon.)	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 geographic controls include distance to the city hall, share of of the parish within the city borders in 1880 and the parish area. The set of topographic controls include the average, maximum and minimum elevations for the parish and the distance to waterways as of 1827. Wealth is inferred from property taxes in 1815, and averaged at the parish level (the dependent variable is the logarithm of the parish average).

Table III. Pollution and shares of low-skilled workers in 1881 – placebo checks with mirror, static and domestic pollution.

Share of low-skilled workers	(1)	(2)	(3)	(4)
Pollution	.0314	.0409	.0314	.0340
	(.0108)	(.0128)	(.0078)	(.0078)
	[.1268]	[.1653]	[.1267]	[.1375]
Static Pollution	.0029			
	(.0093)			
	[.0121]			
Placebo Industry		-.0089		
		(.0123)		
		[-.0360]		
Domestic Pollution			.0133	
			(.0054)	
			[.0539]	
Current Pollution				.0096
				(.0056)
				[.0391]
Observations	4,519	4,519	4,519	4,519
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 the average, maximum and minimum elevations, the distance to waterways as of 1827, the 1817 parish-level shares of farmers, managers and blue-collar workers and latitude/longitude.

Table IV. Pollution and shares of low-skilled workers in 1881 – 2-stage specification.

First stage	Pollution			
	(1)	(2)	(3)	(4)
Synthetic pollution (waterways)	.3239	.3262		
	(.0623)	(.0634)		
Synthetic pollution (steam engines)			.4035	.3979
			(.0948)	(.0943)
Second stage	Share of low-skilled workers (1881)			
	(1)	(2)	(3)	(4)
Pollution	.0613	.0595	.0987	.0961
	(.0171)	(.0169)	(.0298)	(.0302)
	[.2476]	[.2402]	[.3986]	[.3881]
Observations	4,084	4,084	4,084	4,084
F-statistic	27.00	26.45	18.11	17.80
Fixed effects (city)	Yes	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes	Yes
Controls (lat./lon.)	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 cell is the result of a separate regression. The unit of observation is a Lower Super Output Area. The set of extended controls include the average, maximum and minimum elevations, the distance to waterways as of 1827, the 1817 parish-level shares of farmers, managers and blue-collar workers and latitude/longitude. The variable *Synthetic pollution (waterways)* is only constructed for cities with some waterways in 1827.

Table V. 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	.0264	.0301	.0428	.0404	.0410
	(.0051)	(.0078)	(.0092)	(.0086)	(.0084)
	[.2046]	[.2099]	[.2257]	[.2502]	[.2342]
Observations	4,517	4,519	4,519	4,519	4,519
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. 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 the average, maximum and minimum elevations, the distance to waterways as of 1827, the 1817 parish-level shares of farmers, managers and blue-collar workers and latitude/longitude.

Table VI. Selected parameters (see Section 7).

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 VII. Estimated parameters (see Section 7).

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 VIII. 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 7).

Table IX. 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 X. Pollution and shares of low-skilled workers in 1881, 1971, 1981, 1991, 2001 and 2011 in cities with low share of polluted areas versus cities with high share of polluted areas.

Panel A: Low share of polluted areas						
Share of low-skilled	1881	1971	1981	1991	2001	2011
Pollution	.0398	.0311	.0243	.0205	.0188	.0175
	(.0115)	(.0105)	(.0155)	(.0123)	(.0105)	(.0123)
	[.1610]	[.2415]	[.1699]	[.1082]	[.1163]	[.0999]
Observations	2,596	2,595	2,596	2,596	2,596	2,596
Fixed effects (city)	Yes	Yes	Yes	Yes	Yes	Yes
Extended controls	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: High share of polluted areas						
Share of low-skilled	1881	1971	1981	1991	2001	2011
Pollution	.0314	.0330	.0386	.0443	.0472	.0409
	(.0074)	(.0045)	(.0048)	(.0062)	(.0061)	(.0044)
	[.1270]	[.2558]	[.2693]	[.2336]	[.2925]	[.2334]
Observations	1,923	1,922	1,923	1,923	1,923	1,923
Fixed effects (city)	Yes	Yes	Yes	Yes	Yes	Yes
Extended controls	Yes	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. 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 the average, maximum and minimum elevations, the distance to waterways as of 1827, the 1817 parish-level shares of farmers, managers and blue-collar workers and latitude/longitude.

Online Appendix (not for publication)

A 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,

$$\max_{j \in W,E} \left\{ \min_{\ell \in \Omega(j)} A(j,\ell) \right\} < \min_{j \in W,E} \left\{ \max_{\ell \in \Omega(j)} A(j,\ell) \right\}. \tag{12}$$

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 $R^h \leq R^l$ for all $A \leq A^*$. 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,[45]

$$F(A) = \sum_{j \in \{W,E\}} A - \min_{\ell \in \Omega(j)} A(j,\ell), \tag{13}$$

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

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,

$$F(A) = 2(A - d + \rho). \tag{14}$$

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

[45]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 $[\phi(j), \phi(j) + 1]$.

in each neighborhood is,

$$S^l(W) = \gamma - \eta\rho, \tag{15}$$

$$S^l(E) = \gamma + \eta\rho. \tag{16}$$

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. □

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,

$$A^* = \gamma + d(t)/2. \tag{17}$$

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

$$S^l(W,t) = \gamma - d(t)/2, \tag{18}$$

$$S^l(E,t) = \gamma + d(t)/2. \tag{19}$$

Since $d(j,t)$ amenities are persistent, $d(t) > 0$ permanently unless depreciation exists. Equations (18)-(19) then show that sorting persists even after $t = t_c$. □

B Geo-locating 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 *census identifier* i denote a transformation of the book/folio/line numbers in systematic order. 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.

For each entry i, we can define a step function $f : i \mapsto f(i) \in \mathbb{N}$, monotonous in i.[46] The function f defines clusters among census entries.

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:

idi	foliof(i)	blockn(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.

Our true measure of a geographic cluster (i.e. 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

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

and we will discuss sensitivity analyses 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 geo-referencing a census address incorrectly by limiting the pool of potential matches (the contemporary geo-located addresses) 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.

For our geo-localization algorithm, 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 geo-locate the remaining households.

Recognizing clusters and inference We infer the geo-location of all households of the same census cluster f from the geo-location 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. Geo-locate all geo-referenced households.

2. Divide a parish into 4 equal regions, depending on their position relative to the maximum and minimum latitudes and the the maximum and minimum 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 and newly selected observations, and re-iterate.

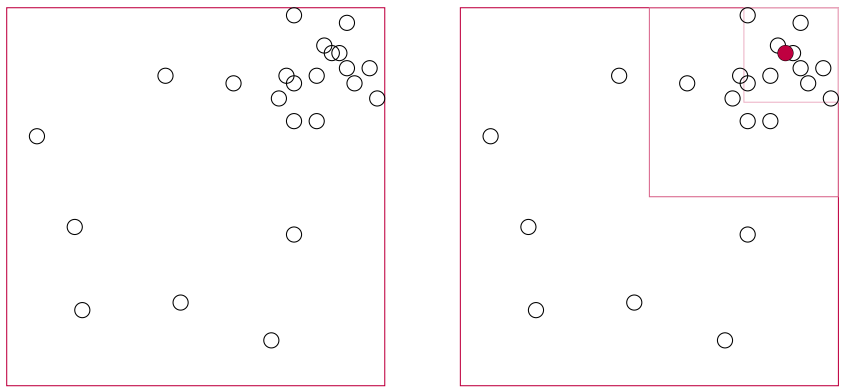
5. Stop after a given number of iterations, generate the average latitude and longitude among the remaining households, and attribute this geo-reference to *all* households in the same census cluster f.

A graphical illustration of this algorithm is provided in Figure A1 with 2 iterations. Two (resp. three) iterations already divide a parish into 16 (resp. 64) small regions. Note that we can always generate a dispersion of geo-references at each step of the algorithm and keep track of problematic situations, e.g., the existence of two separate and equally numerous clusters.

The advantage of this process is not only to infer geo-references for unmatched households but also to smooth geo-references among already-matched units.

We then overlay these newly-identified blocks with our consistent geographic units (LSOAs) and attribute a unique LSOA identifier to all households in the 1881 Census.

Figure A1. Finding geographic clusters among geo-referenced 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. 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 many different census clusters (grouping 1, 2, 5 or 10 folios 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.

C Additional evidence on the nature of neighborhood effects

The theoretical analysis developed in Section 7 is silent about the 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 Appendix provides a large set of descriptive statistics about formerly polluted neighborhoods in 2011. While this analysis is not causal and cannot be used as hard evidence in favor of one particular channel of persistence, it helps understand the within-city distribution of consumptive amenities and its relationship with past atmospheric pollution.

We use various data sources to construct indicators of consumptive amenities at the 2001 LSOA resolution. First, we compute the density of different public services (schools and universities, theaters, museums and libraries, parks, churches and hospitals) in 1881 and 2011. In 1881, we digitize this information from the same set of 1880-1900 Ordnances Survey maps that we use to detect chimneys. In 2011, we use the Point of Interest (POI) data provides by the Ordinance Survey to determine the local provision of amenities, i.e. public transport, health, education and leisure services.[47] Second, we collect a snapshot of housing characteristics: house ages from the Consumer Data Research Centre and house characteristics from transaction data (Land Registry and Nationwide). Third, 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 (student test score).[48] Fourth, 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.[49] Finally, we collect sub-indices of the 2010 English Indices of Deprivation (Income; Employment; Health and Disability; Education,

[47]In 1880-1900, we observe schools, hospitals, parks or public administration. In 2011, we add to the previous list: local and national government buildings, courts and police stations, bus or train stations, botanical gardens and zoos.

[48]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.

[49]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.

Skills and Training; Barriers to Housing and Services; Crime; Living Environment).[50]

We then 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 Table A5), 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 all 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–as also shown by the prevalence of social housing in Table A7: 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. There is also a small difference in the number of bedrooms or square meters across neighborhoods.

[50]The English Indices of Deprivation (2010), The Social Disadvantage Research Centre at the Department of Social Policy and Social Work, University of Oxford.

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 the estimates are quite small in magnitude. The demand for high-quality amenities in good neighborhoods may be counteracted by high land prices.[51] Note that we do not observe a systematic correlation between pollution and public amenities in 1881 (unreported tests).

In a second specification (see Table A8), 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 A5. By contrast, half of the correlation between neighborhood composition and past pollution is captured by social housing and an additional quarter disappears once we control for other observable indicators of 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 indicators of schooling supply and the presence of police forces do not seem to alter the benchmark relationship, adding indicators of school composition and crime incidence captures between 20 and 40% of the initial relationship.

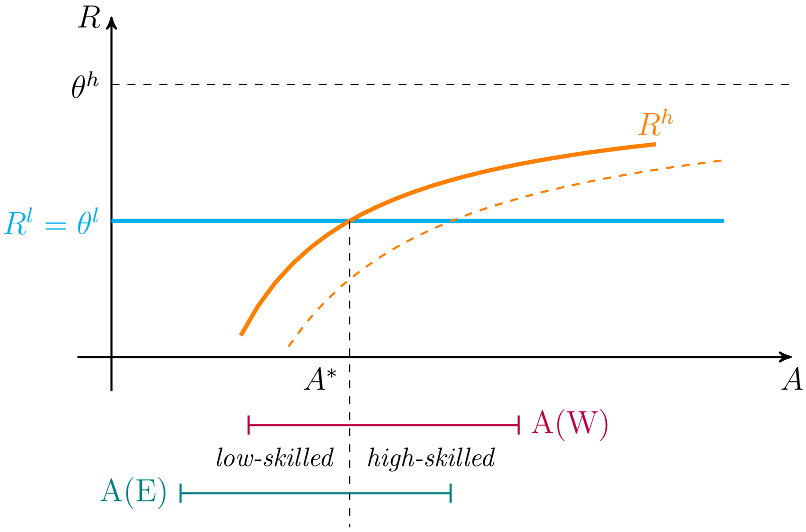
These descriptive statistics indicate that the persistence in segregation does not relate to rigid consumptive amenities (e.g., provision of public services, Victorian housing stock or old private schools constructed during industrialization) which would anchor neighborhoods in a certain equilibrium 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 (information on consumptive amenities would be needed along the whole process of segregation), such investigation would be essential to derive policy implications that rely on targeting the provision of these amenities instead of the initial distribution

[51]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.

of pollution.

D Additional figures and tables

Figure A2. Amenities and Neighborhood Sorting: Equilibrium with pollution and $\gamma = \frac{1}{2}$.

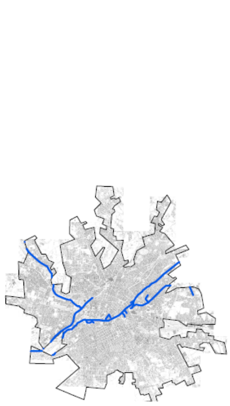


Notes: the x-axis represents the level of consumptive amenities and the y-axis is the rent. $A(W)$ and $A(E)$ depict the distribution of amenities in the two neighborhoods $\{E, W\}$. We assume that the amenity levels overlap across neighborhoods even with pollution and we focus on the cases where a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

Figure A3. Town maps – marking and identifying chimneys.

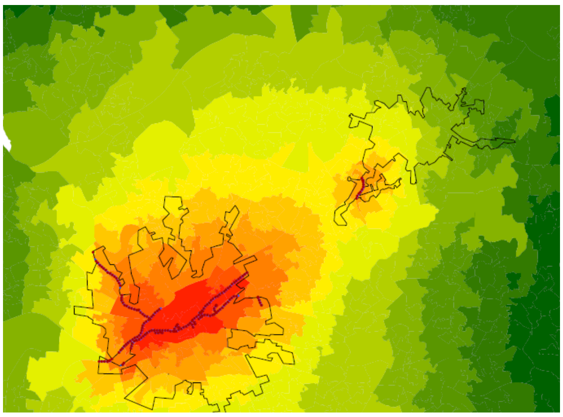
Sources: Ordnance Survey Maps - 25 inch to the mile, 1842-1952. Marks *X* and the identifiers, e.g., *00006*, are used by a recognition algorithm to locate chimneys and associate a factory.

Figure A4. An illustration of the 2-stage empirical approach in Manchester (synthetic chimneys located along 1827 waterways).

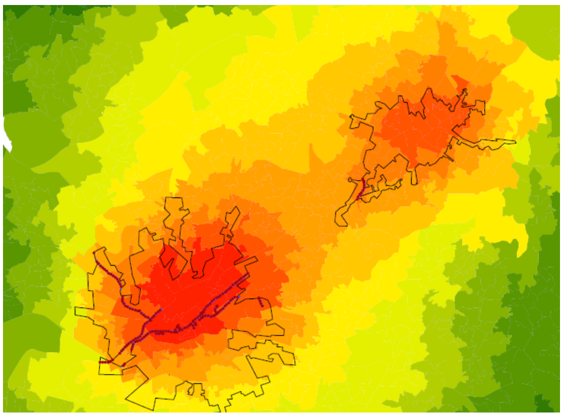


(a) City Map

(b) Locating chimneys along waterways.



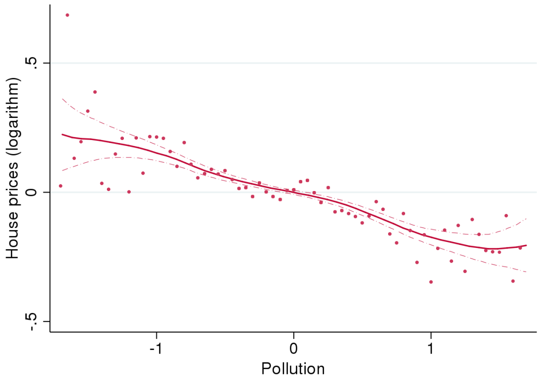
(c) Synthetic pollution.



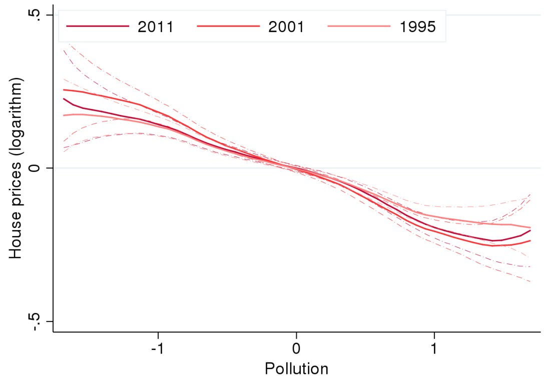
(d) Actual pollution.

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 A5. House transaction prices (y-axis) and pollution (x-axis) across neighborhoods – average and evolution between 1995 and 2011.



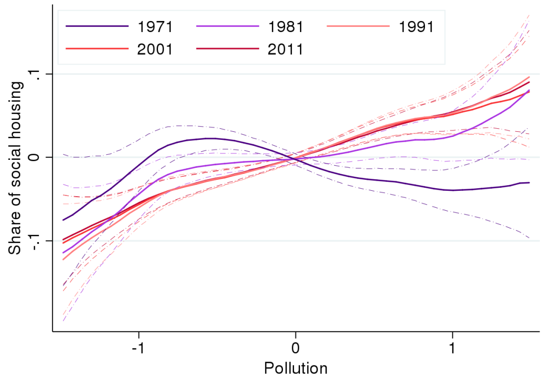
(a) Average



(b) Evolution

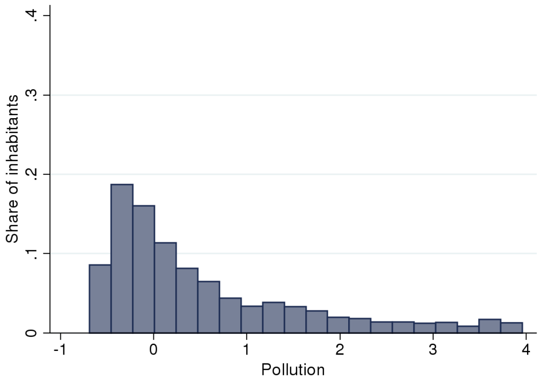
Notes: The left (resp. right) panel represents the relationship between the (logarithm of the) average transaction prices between 2000 and 2011 (resp. in 1995, 2000, and 2011) and our (standardized) measure of past pollution. We consider the residuals of all measures once cleaned by city Fixed-Effects, 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 A6. Social housing (y-axis) and pollution (x-axis) across neighborhoods in 1971, 1981, 1991, 2001 and 2011.

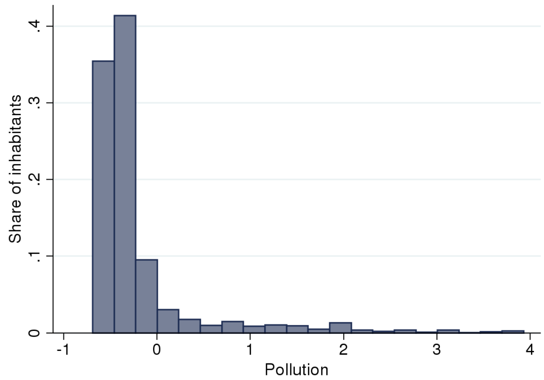


Notes: The figure represents the locally weighted regressions on all observations between the shares of social housing and our (standardized) measure of past pollution. We consider the residuals of all measures once cleaned by city Fixed-Effects, geographic and topographic controls.

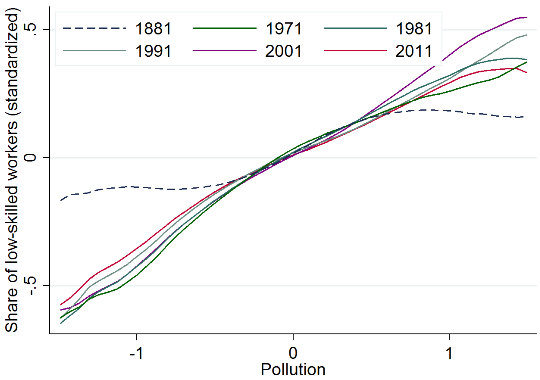
Figure A7. Importance of the city-wide distribution of deprivation and pollution in sorting across neighborhoods.



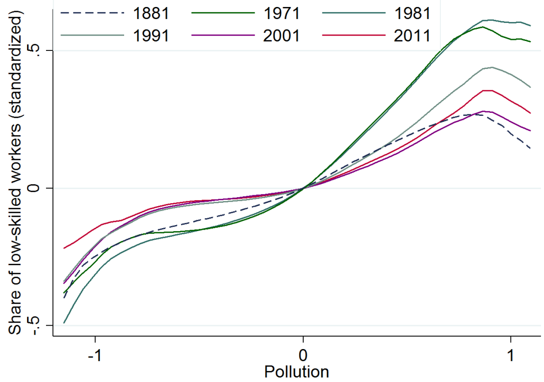
(a) Distribution (high share).



(b) Distribution (low share).



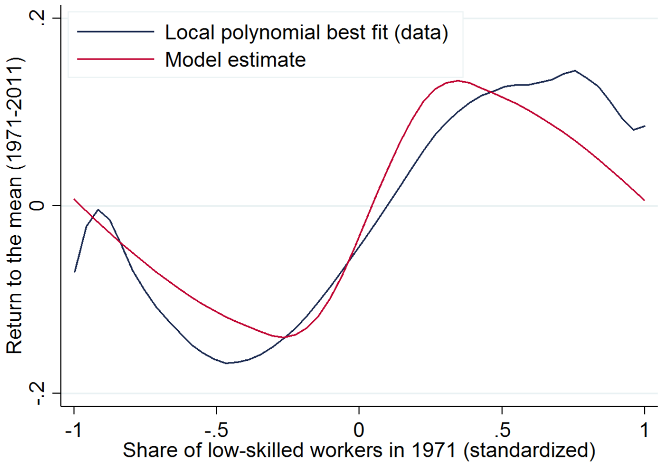
(c) Persistence (high share).



(d) Persistence (low share).

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 panel (resp. right panel), we keep cities for which there is a low (resp. high) share of polluted areas compared to the share of low-skilled workers, i.e., we keep the observations with $share_{polluted}$ above the median (resp. below the median). $share_{polluted}$ is the share of areas with pollution above the mean. 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 shares of low-skilled workers within a pollution-bin. The lines are locally weighted regressions on the observed sample.

Figure A8. 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 geographic and topographic controls.

Table A1. Descriptive statistics and variance decomposition.

VARIABLES	Obs.	Mean	Standard deviation total	between	within
		Air pollution			
Normalized pollution	4,524	0	1	.542	.774
		Segregation measures (shares)			
*1817**					
Low-skilled workers	4,524	.782	.113	.077	.074
High-skilled workers	4,524	.128	.099	.067	.053
Farmers	4,524	.088	.087	.065	.068
1881					
Low-skilled workers	4,524	.607	.247	.153	.225
High-skilled workers	4,524	.282	.175	.130	.207
Farmers	4,524	.111	.193	.171	.167
2011					
Low-skilled workers	4,524	.583	.175	.121	.119
High-skilled workers	4,524	.416	.175	.121	.119
		City controls			
Number of chimneys	4,524	.682	3.47	.938	3.34
Distance town hall (m)	4,524	4823	5334	4754	1487
Distance parks (m)	4,524	9041	22461	27362	1153
Share LSOA within city borders	4,524	.356	.435	.269	.299
Area (square km)	4,524	1.64	6.72	7.24	5.69
		Topographic controls			
Maximum elevation (m)	4,519	72.3	67.4	69.4	34.4
Minimum elevation (m)	4,519	50.2	47.8	44.8	18.55
Mean elevation (m)	4,519	60.5	55.0	54.3	23.3
Distance waterways (m)	4,524	5723	14380	17899	1391

Notes: * Shares in 1817 are computed at the parish-level, which explains the lower variance.

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

Share of low-skilled workers	1881	2011
Pollution	3.014	.0410
	(.0076)	(.0084)
Distance hall (inverse)	3.014	-2.387
	(3.011)	(2.851)
Distance waterways (inverse)	.0717	.0300
	(.0155)	(.0353)
Share area (city)	.0138	-.0022
	(.0149)	(.0120)
Distance (city borders)	.0013	.0148
	(.0067)	(.0049)
Distance (parks)	.0041	.0066
	(.0041)	(.0035)
Area	-.0027	-.0012
	(.0007)	(.005)
Maximum elevation	.0001	-.0004
	(.0004)	(.0002)
Minimum elevation	-.0012	.0000
	(.0004)	(.0004)
Average elevation	.0002	-.0005
	(.0007)	(.0004)
Number of chimneys	.0700	-.0013
	(.0152)	(.0005)
Distance to chimney (textile)	-.0037	.0006
	(.0033)	(.0024)
Distance to chimney (chemical)	.0023	-.0010
	(.0026)	(.0022)
Share low-skills (1817)	.1126	.0633
	(.0729)	(.0731)
Share farmers (1817)	-.1587	-.0480
	(.0764)	(.0774)
Longitude	.0038	.0077
	(.0021)	(.0022)
Latitude	.0002	.0001
	(.0024)	(.0018)
Observations	4,519	4,519

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 workers are defined as manual unskilled and semi-skilled workers, and job seekers. Managers, rentiers, clerks, manual skilled workers and farmers are not included. The average share of low-skilled workers in 1881 is .61.

Table A3. Pollution and shares of low-skilled workers – difference-in-difference specifications.

Panel A: LSOA, 1817-1881				
Share of low-skilled workers	(1)	(2)	(3)	(4)
Pollution	.0355	.0338	.0320	.0316
	(.0057)	(.0058)	(.0063)	(.0064)
	[.1863]	[.1774]	[.1681]	[.1662]
Observations	8,696	8,696	8,696	8,696
Fixed effects (LSOA)	Yes	Yes	Yes	Yes
City×Year FEs	Yes	Yes	Yes	Yes
Geography×Year FEs	No	Yes	Yes	Yes
Topography×Year FEs	No	No	Yes	Yes
Coordinates×Year FEs	No	No	No	Yes
Panel B: parish, 1817-1881				
Share of low-skilled workers	(1)	(2)	(3)	(4)
Pollution	.0535	.0520	.0350	.0332
	(.0171)	(.0173)	(.0192)	(.0194)
	[.3076]	[.2993]	[.2012]	[.1913]
Observations	1,034	1,034	1,034	1,034
Fixed effects (parish)	Yes	Yes	Yes	Yes
City×Year FEs	Yes	Yes	Yes	Yes
Geography×Year FEs	No	Yes	Yes	Yes
Topography×Year FEs	No	No	Yes	Yes
Coordinates×Year FEs	No	No	No	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/year. In Panel A (resp. Panel B), the basic specification is a panel specification with LSOA (resp. parish) fixed-effects and city-specific trends. The set of geographic controls include distance to the city hall, share of LSOA (resp. parish) within the city borders in 1890 and the area for the LSOA (resp. parish). The set of topographic controls include the average, maximum and minimum elevations for the LSOA (resp. parish) and the distance to 1827 natural waterways.

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	.0391	.0364	.0303
	(.0079)	(.0088)	(.0090)
	[.1580]	[.1471]	[.1225]
Observations	4,519	4,519	4,519
Fixed effects	Parish	Ward	MSOA
Panel B: Clusters	Share of low-skilled workers (1881)		
	(1)	(2)	(3)
Pollution	.0350	.0350	.0350
	(.0052)	(.0057)	(.0076)
	[.1415]	[.1415]	[.1415]
Observations	4,519	4,519	4,519
Clusters	MSOA	Ward	City
Panel C: Sample	Share of low-skilled workers (1881)		
	(1)	(2)	(3)
Pollution	.0329	.0558	.0358
	(.0061)	(.0116)	(.0064)
	[.1328]	[.2255]	[.1447]
Observations	3,056	3,533	4,285
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 topographic controls include the average, maximum and minimum elevations for the LSOA and the distance to 1827 natural waterways . The set of 1817 controls include the parish-level shares of farmers, managers and blue-collar workers. Low-skilled workers are defined as manual workers and employees (categories 4 to 9 in the 2011 Census – see section 4). A *MSOA* (Medium Super Output Area) is the second smallest unit in the census, and there are 1600 MSOAs in our sample. A *ward* is an electoral ward (election for local councils): there are 1200 wards in our sample. *London* is Greater London and thus include 33 districts in addition to the City of London. *NW* is the North-Western region while *NE* is the North-Eastern region.

Table A5. 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	.0640	.0701	.0539	.0787	.0338	.0034	.0312	.0394
	(.0138)	(.0164)	(.0140)	(.0135)	(.0107)	(.0053)	(.0068)	(.0146)
	[.2501]	[.2465]	[.1928]	[.2699]	[.1292]	[.0113]	[.1263]	[.1657]
Observations	4,519	4,519	4,519	4,519	4,519	4,519	4,519	4,519
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	-.0050	-.0038	.0072	-.0000	.0038	.0428	.0123	.0645
	(.0027)	(.0012)	(.0012)	(.0001)	(.0047)	(.0101)	(.0021)	(.0126)
	[-.0371]	[-.0791]	[.0826]	[-.0025]	[.0287]	[.1225]	[.1561]	[.1796]
Observations	4,519	4,519	4,519	4,519	4,519	4,519	4,519	4,519
Panel C: Housing quality	Building 1900	Building 1970	Building 2000	Year of construction	Square meters	Bedrooms	Flats	Detached
Pollution	-.0164	-.0078	.0118	2.287	-1.832	-.0140	.0550	-.0227
	(.0104)	(.0109)	(.0060)	(1.719)	(.7066)	(.0170)	(.0138)	(.0056)
	[-.0638]	[-.0372]	[.0678]	[.0668]	[-.0703]	[-.0220]	[.1996]	[-.1481]
Observations	4,519	4,519	4,519	4,228	4,228	4,228	4,519	4,519
Panel D: Amenities	Parks	Entert.	Church	Hospital	Public	Justice	Transport	Botanical
Pollution	.0401	-.0011	.0180	-.0050	.0322	.0116	.0361	-.0090
	(.0169)	(.0204)	(.0095)	(.0023)	(.0203)	(.0093)	(.0118)	(.0041)
	[.0735]	[-.0016]	[.0571]	[-.0498]	[.0653]	[.0559]	[.1181]	[-.0401]
Observations	4,519	4,519	4,519	4,519	4,519	4,519	4,519	4,519

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 controls correspond to the ones used in column 5 of Table I. 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 4). 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 registry of transactions. Panel D uses as dependent variables the number of Parks and recreation areas, Theaters, museums, Churches, Hospitals, National and local authorities, Justice (courts and police stations), Transport (bus and train stations), Botanical gardens and zoos per 100 inhabitants in 2011.

Table A6. 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	-.1042	-.0801	-.1067	-.0513	-.0781	-.1515
	(.0190)	(.0147)	(.0185)	(.0109)	(.0226)	(.0248)
	[-.1668]	[-.1282]	[-.1888]	[-.0908]	[-.1421]	[-.2757]
Observations	4,519	4,519	4,519	4,519	4,519	4,519
Fixed effects (city)	Yes	Yes	Yes	Yes	Yes	Yes
Controls (house ch.)	No	Yes	No	Yes	No	Yes
Controls (topography)	Yes	Yes	Yes	Yes	Yes	Yes
Controls (1817)	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors are reported between parentheses and are clustered at the parish-level. Each column is the result of a separate regression. The unit of observation is a Lower Super Output Area. 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 columns 1 and 2, controls for house types 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 3 to 6, controls for house types include the average shares of detached, semi-detached, terraced houses and new houses for all transactions.

Table A7. Pollution and social housing/migrant shares (1971-2011).

Effect of pollution on ...	1971	1981	1991	2001	2011
Social housing	.0089	.0504	.0659	.0597	.0572
	(.0072)	(.0099)	(.0084)	(.0071)	(.0073)
	.287	*.358*	*.297*	*.260*	*.232*
Owners	-.0413	-.0543	-.0694	-.0720	-.0740
	(.0072)	(.0083)	(.0085)	(.0086)	(.0095)
	.429	*.494*	*.580*	*.583*	*.535*
Migrants (New Commonwealth)	.0129	.0189	.0173	.0195	.0307
	(.0034)	(.0046)	(.0046)	(.0054)	(.0073)
	.041	*.060*	*.064*	*.085*	*.128*
Migrants (Other)	.0012	.0008	.0006	.0028	.0061
	(.0010)	(.0008)	(.0008)	(.0009)	(.0013)
	.034	*.035*	*.043*	*.053*	*.075*
Observations	4,517	4,519	4,519	4,519	4,519
Fixed effects (city)	Yes	Yes	Yes	Yes	Yes

Standard errors are reported between parentheses and are clustered at the parish-level. 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 controls include geographic coordinates, the average, maximum and minimum elevations for the LSOA and the distance to waterways as of 1827. Social housing (resp. Owners) are the shares of households in a social housing (resp. owners) as captured in the Census (the unit of calculation is the 2011 LSOA).

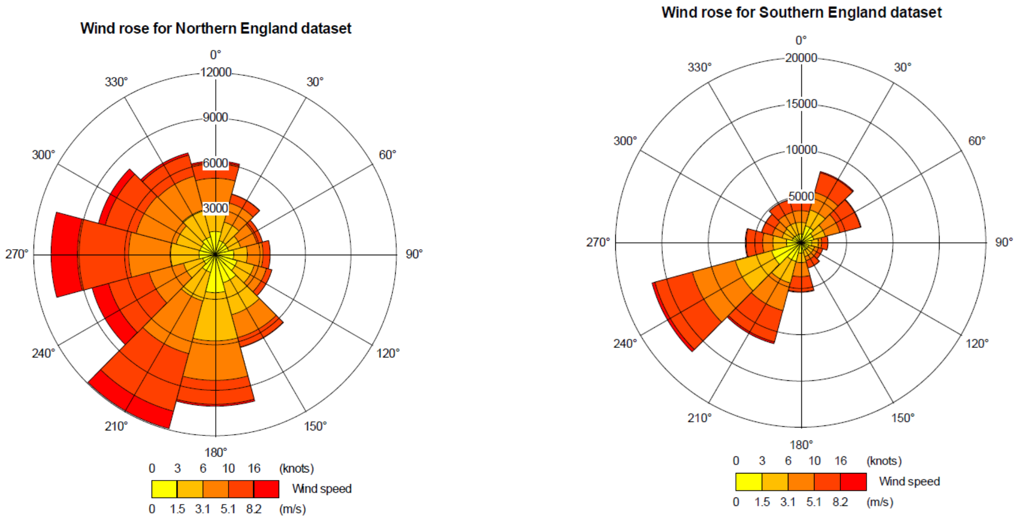
Table A8. Pollution and shares of low-skilled workers – controlling for amenities, housing characteristics and crime and education indicators.

Panel A: Amenities	1881		2011	
Share of low-skilled workers	(1)	(2)	(3)	(4)
Pollution	.0309	.0285	.0339	.0280
	(.0065)	(.0066)	(.0071)	(.0065)
	[.1250]	[.1153]	[.1933]	[.1599]
Observations	3,814	3,814	3,814	3,814
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	.0545	.0380	.0203	.0114
	(.0102)	(.0080)	(.0056)	(.0041)
	[.3109]	[.2167]	[.1161]	[.0650]
Observations	995	4,228	4,228	4,228
Sample	New housing	All	All	All
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	.0313	.0219	.0356	.0276
	(.0059)	(.0056)	(.0070)	(.0057)
	[.1788]	[.1251]	[.1933]	[.1599]
Observations	4,519	1,792	4,519	4,519
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. Each column is the result of a separate regression. The unit of observation is a Lower Super Output Area. The basic set of controls correspond to the ones used in column 5 of Table I. 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.

E Atmospheric Dispersion Model

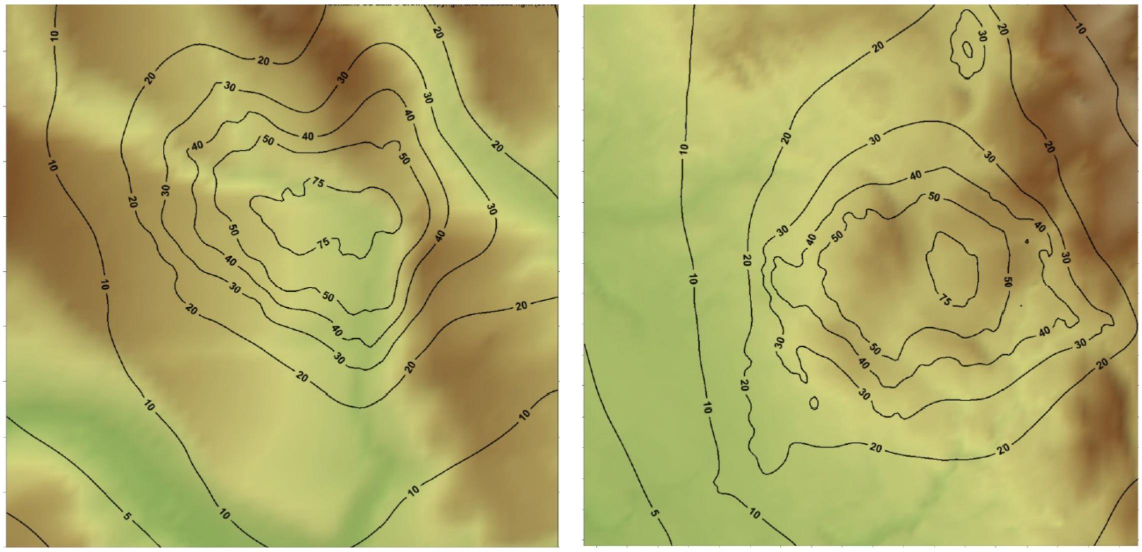
Figure A9. Wind roses differences across two sets of meteorological conditions.



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

Sources: Met Office – 10-year statistical meteorological data. We use 5 different sets of meteorological conditions across England that we associate to our 70 metropolitan areas.

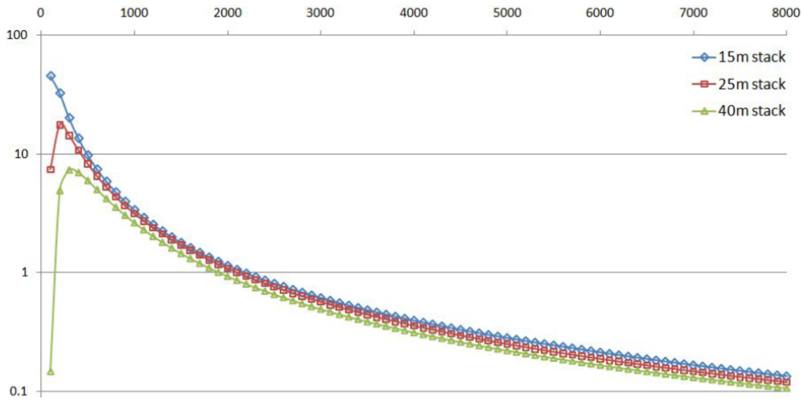
Figure A10. Topography and industrial air pollution.



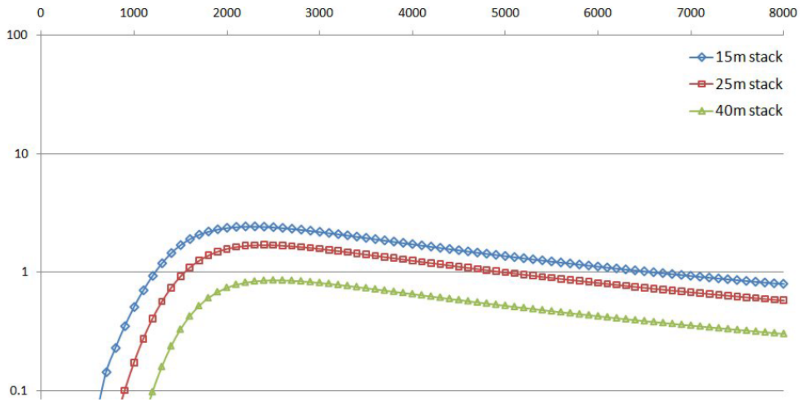
(a) Halifax. (b) Olham.

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

Figure A11. Wind patterns and industrial air pollution – stable and unstable conditions.



(a) Unstable conditions.

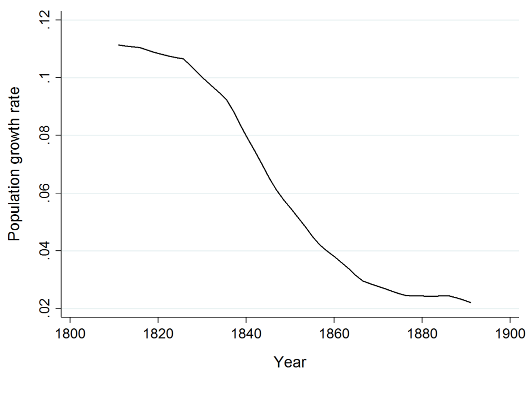


(b) Stable conditions.

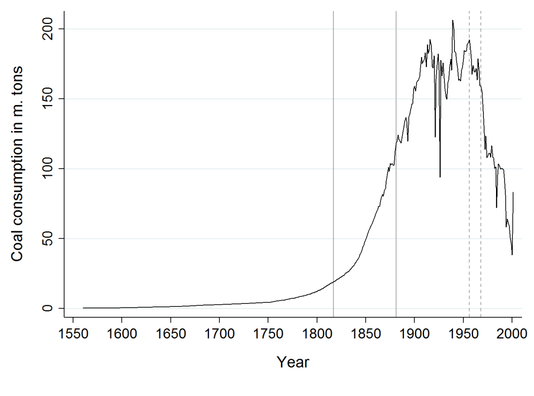
Sources: ADMS 5. The y-axis is the ground-level concentration for 1g/s emission rate ($\mu g/m^3$, logarithmic scale), and the x-axis is the downwind distance (m).

F Descriptive statistics

Figure A12. Migration and coal consumption during the Industrial Revolution.



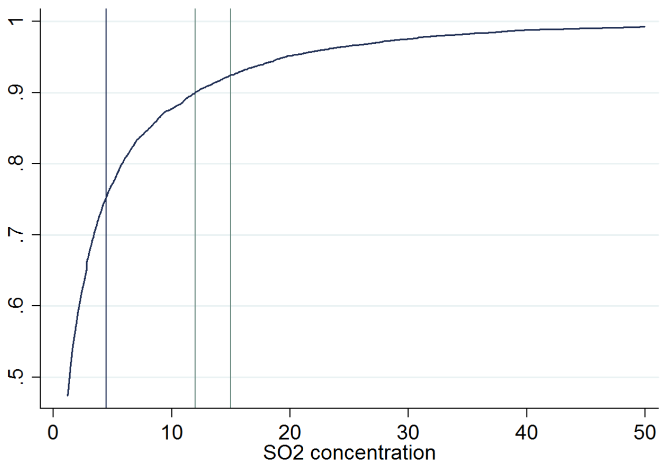
(a) Population Growth Rates in cities (1801-1891)



(b) Coal consumption in million tons (1560–2001)

Notes: The left panel plots the average decadal population growth rate for the period 1801-1891 in our sample cities. The right 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 two solid grey lines indicate the years 1817 and 1881 for which we have detailed occupational information within cities. The dashed grey lines mark the introduction of the 1956 Clean Air Act and the stricter 1968 Clean Air Act. Sources: Warde, 2007.

Figure A13. Cumulative of pollution in our sample of 10000 parishes and National Ambient Air Quality Standards (12-15 $\mu g/m^3$).



Sources: Authors' calculations.

Table A9. 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 and authors' calculation.

Figure A14. Cities in our sample.