East Side Story: Historical Pollution and Persistent Neighborhood Sorting*

Stephan Heblich Alex Trew Yanos Zylberberg

August 9, 2016

Draft – comments welcome

Abstract

Why are the East sides of former industrial cities like London or New York poorer and more deprived? We argue that this is the result of prevailing wind patterns causing higher concentrations of industrial pollution in the East side of these cities. To capture this empirically, we geolocate nearly 5,000 industrial chimneys in 70 English cities around the year 1880 and use an atmospheric dispersion model to predict where their smoke would have drifted. Individual-level census data for 1881 show that pollution induced neighbourhood sorting with the working class population residing in the East. These equilibria persist to this day. Historical pollution patterns explain up to 20% of modern within-city deprivation even though the pollution that initially caused it has now waned. A quantitative model shows the role of non-linearities and tipping-like behaviors in such persistence.

Keywords: Neighborhood Sorting, Historical Pollution, Deprivation, Persistence, Wind Direction.

JEL codes: R23, Q53, N90

*Heblich: Bristol University; stephan.heblich@bristol.ac.uk; Trew: University of St. Andrews; awt2@st-andrews.ac.uk; Zylberberg: Bristol University; yanos.zylberberg@bristol.ac.uk. We are grateful to Nate Baum-Snow, Kristian Behrens, Davide Cantoni, Klaus Desmet, Oliver Falck, Walker Hanlon, Robert McMillan, Giacomo Ponzetto, Steve Redding, Jorge De la Roca, Max Satchell, Hannes Schwandt, William Strange, Leigh Shaw-Taylor, Ludgar Woessmann for useful discussions and comments. We also thank seminar participants in Bristol, Cambridge, CERGE-EI, IEB Workshop on Urban Economics, Ifo Institute, Munich, SERC, St Andrews, Toronto, and Warwick for helpful comments. We are grateful to Nicholas Cheras, Qingli Fang, Joanna Kalemba, Matthew Litherland, Filip Nemecek, Ondrej Ptacek for excellent research assistance. 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*

In de-industrialising economies, cities that were formerly reliant on industry tend to have Eastern suburbs that are notably poorer than the Western suburbs. This observation is echoed in media stories about the East Side in London, New York, Paris or Vancouver 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 rationalize this observation as follows. Neighborhoods to the East are poor because they were most affected by highly polluting factories that emerged during the Industrial Revolution in the nineteenth century. The prevailing wind patterns in the UK, as well as Western Europe and the US, are called *Westerlies* – they blow from the West to the East. Historically, factories were typically located in the center of town so the winds created asymmetric pollution effects with the East being more polluted than the West. The consequently uneven distribution of amenities, in the form of clean air, induced a neighborhood sorting process. The middle class and upper class moved to the less polluted western part of town while the working class had their residential areas in the East. This sorting turns out to be surprisingly stable and the effects of the now absent pollution are still felt to the modern day: Almost 50 years after the drastic reduction of coal burning in these cities, the former concentrations of pollution from historical factories explain 20% of current deprivation.

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 this dataset. First, we develop an algorithm to determine the coordinates of all industrial chimneys from historical Ordnance Survey (OS) maps of the 70 largest metropolitan

areas in England.[1] These highly detailed maps were published over the period 1880-1900. Second, we use pollution modelling software (ADMS 5)[2] that incorporates within-city information on terrain, wind directions, sector-specific coal burning intensity, chimney dimensions, exit velocity and coal burning temperature to predict the pollution from each individual chimney. To aggregate pollution across individual chimneys, we superimpose 50×50 meter grid cells over each city and sum the chimney-specific pollution concentration in every grid cell.[3] Third, we develop a novel algorithm to geo-locate individuals in the 1881 Census into very small administrative units, for our purpose 2001 Lower Super Output Areas (LSOA).[4] This is because existing data from the 1881 censuses are of limited use for *within-city* analysis since the data are nested in comparatively large spatial units, i.e., parishes. For instance, across the 70 metropolitan areas in our sample, we observe about 4,500 LSOAs but only 500 parishes and cities like Bristol, Liverpool or Manchester are covered by about 90 LSOAs but only 10 to 20 parishes.

We find that an increase of one standard deviation in air pollution is associated with an additional 0.15 standard deviations in the share of low-skilled workers in 1881. A differential in pollution 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. This effect is robust to the addition of a large set of controls including distance to the major public amenities in the city, distance to canals, neighborhood composition in 1817, latitude and longitude, and fixed effects at the parish or Medium Super Output Areas (thus controlling for potentially correlated East/West differences).

The ideal experiment to identify the impact of pollution on neighborhood sorting would be to randomly locate a chimney, and compare upwind and downwind neighborhoods at the same distance from the chimney. To get closer to this thought experiment, we first include proxies capturing the proximity to factories. However, conditioning on distance does not take into consideration that chimneys may have been selectively located upwind of poor areas. To account for reverse causality,

[1] Unless stated otherwise, chimney always refers to factory/industrial chimneys (or smoke stacks). We will also consider domestic chimneys but their contribution to overall pollution is small during the heyday of the industrialization.

[2] Atmospheric Dispersion Modelling System (ADMS) models have been developed to make use of the most up-to-date understanding of the behaviour of the lower levels of the atmosphere. ADMS 5 is the world leading modelling system for atmospheric emissions.

[3] We validate our pollution measure with some external sources: we use a report from the Air Pollution Advisory Board in 1915, and find strikingly similar patterns between our measure and the reported soot deposits at ten locations around Manchester.

[4] The average LSOA in our sample covers an area of 1.2 square kilometers and the median is 0.3 square kilometers.

we run a series of robustness checks. Specifically, we discard the existence of unobserved pre-existing (dis)amenities in polluted neighborhoods by looking at the neighborhood composition in 1817, and by running difference-in-difference specifications between 1881 and 1817. To reduce concerns about potential changes in (dis-)amenities between 1817 and 1881, we instrument the pollution pattern induced by *actual chimneys* with two *counterfactual pollution patterns*. We first locate counterfactual chimneys along natural waterways and canals present in 1827 within the cities, and create the associated synthetic pollution dissemination. The rationale is that steam engines needed water for cooling and canals were the best available water source in the absence of effective municipal water supplies (Maw et al. (2012)). Note that, as we condition on distance to the canal, this specification exploits the difference between upwind and downwind neighborhoods at the same distance from potential factories located along waterways. Our second counterfactual pollution pattern relies on a random (and uniform) allocation of chimneys within 1890 city borders. This instrument draws exclusively from variation induced by wind patterns and topography but not industry location. The 2-stage specifications with both instruments deliver similar qualitative results, but 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 historic pollution on the social composition of neighborhoods. The estimates in 1971, 1981, 1991, 2001 or 2011 are 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, by house prices or official deprivation indices. Interestingly, the dynamics of persistence show patterns of non-linearities, with some mean-reversion for intermediate values of within-city pollution and a strong inertia for neighborhoods with extreme values of within-city pollution.

In order to characterize the non-linearities in the persistence, we develop a stylized model of neighborhood sorting that expands on that in Lee and Lin (2013). The location choices of high- and low-income individuals are a function of consumptive amenities with some amenities being endogenously tied to the neighborhood composition (e.g., through tipping-like preferences, public goods accumulation or other man-made amenities). We estimate the model to match the dynamics over

[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 Figure A3, the second Clean Air Act in 1968 caused a much more abrupt drop in coal consumption.

the period 1881-1971, and find that the best fit exhibits tipping-like behaviors with large tail effects for the endogenous consumptive 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. An 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.

Our findings are relevant for contemporary policy issues. First, the long-run impact of environmental disamenities on the spatial organisation of cities holds important implications for the design of environmental and urban policies in industrial countries like China, or economies in the process of structural transformation. We show that extreme pollution concentrations are associated with extreme neighborhood sorting. Second, many developed economies rely on costly urban policies to promote social diversity and open up deprived areas to new housing opportunities and business investment.[6] We identify a tipping force in the dynamics of segregation which is useful for policy makers who want to prevent a rise in spatial inequalities.

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 modified the spatial organization of cities in the long-run, with the East Sides being permanently more deprived. 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). Our argument further relates to Depro et al. (2015) who argue that neighborhood sorting, rather than environmental injustice, is the cause of the observation that poor households disproportionately 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). Even after the sharp decrease of industrial pollution in English cities, East Side neighborhoods remain the poorer part of town. Our quantitative analysis points to non-linearities and tipping-like behavior as the main driver of the dynamics of segregation: past a certain threshold, highly-polluted neighborhoods accumulate low consumptive ameni-

[6] See Neumark and Simpson (2015) for a discussion of place-based policies.

ties and attract low-income residents even after pollution has waned. We differ from most papers in this literature in two dimensions. In our context with fewer ethnic tensions, we mostly identify a social component behind segregation (in contrast to the literature, with a strong focus on the United States and ethnic considerations). Besides, we exploit a temporary disamenity to explain the initial spatial distribution of residents, thereby cleaning for differences in long-term amenities across neighborhoods (as in Lee and Lin, 2013, for instance).

The phenomenon of segregation relates closely to the phenomenon of gentrification which studies the current rise of historic centers in the United States (Brueckner and Rosenthal, 2009; Brueckner and Helsley, 2011; Guerrieri et al., 2013). Our initial distribution of deprivation may induce a similar revitalization of East Side neighborhoods in formerly-polluted English cities. While there exist striking examples (Canary Wharf in London or the quays in Manchester and Liverpool), we only find signs of reversion to the mean for neighborhoods with initially moderate levels of pollution.

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 a 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, we do not provide theoretical micro-foundations and cannot untangle them in the empirical analysis.

Fourth, our paper relates to research on the emergence and persistence of 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 (Industrial Revolution in England), 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 that operate through residential sorting. Another related paper is Redding and Sturm (2016), who use Second World War destruction in London to identify patterns of spatial sorting accross neighborhoods.

Fifth, we make several methodological contributions to quantitative research in economic history.[7] Our first contribution is to provide a methodology to digitize

[7]Our use of exogenous variation in weather conditions to study an economic outcome also relates to the new climate-economy literature (Dell et al., 2014).

historical maps and fully exploit them as extremely valuable and detailed sources of information. Related to this approach is work by Hornbeck (2012) 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 past 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 the effects of pollution exposure on health. We do not attempt to summarize the literature here and instead refer interested readers to Graff Zivin and Neidell (2013) for a recent overview. Of this comprehensive body of literature, our paper is most closely related to historical assessments of the effect of coal use on health. Recent contributions are Beach and Hanlon (2016) which looks at mortality effects of coal pollution in the UK during the period 1851–1911; Clay and Troesken (2011), who look at the London fog; and, Barreca et al. (2014) and Clay et al. (2016), who estimate the effects of coal consumption on mortality rates in the U.S. during the mid 20th century. Other recent studies of the effect of pollution on non-health outcomes include those on violent crime, such as Herrnstadt and Muehlegger (2015) and Heyes and Saberian (2015).

The remainder of the paper is organized as follows. In Section 2, we develop a stylized model of neighborhood sorting and analyze the impact of pollution on city segregation. Section 3 briefly provides some elements of context. We detail our main data sources and our methodology to construct a pollution imprint and geo-locate individuals in the past census in Section 4. Section 5 then outlines our empirical strategy. 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. Section 8 concludes with implications for pollution reduction policies and neighborhood revival policies today.

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 version that we further develop in Section 7.

A city is composed of two neighborhoods, West and East, and pollution affects neighborhoods differently because of prevailing wind patterns. The difference in air quality causes sorting of a portion of the high- (low-)skilled workers into the less (more) polluted neighborhood. As in Lee and Lin (2013), pollution acts as a 'natural amenity' and causes an anchoring of neighborhoods. We abstract from the production side of the economy,[8] such that sorting arises out of within-city differences in consumptive amenities.

2.1 Environment

A city is composed of two neighborhoods indexed $j \in \{W, E\}$ (West and East) within which firms are uniformly distributed at locations $\ell \in \Omega(j)$. 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 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. In contrast, the location factor varies within a neighborhood – different locations within a neighborhood have inherent differences in attractiveness (see Davis and Dingel, 2014).[9] We assume that $x(\ell, j)$ is, in both

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

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

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

2.2 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. We normalize $\bar{V}^l = 0$, so that 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 $I^l(j, \ell) = 1$ 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. □

2.3 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 $2A = 2\gamma$,

$$A^* = \gamma. \tag{6}$$

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, each neighborhood has the same $\min_\ell\{A(j,\ell,t)\} = 0$, so $S^l(j,t) = \gamma$ 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:

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

While both neighborhoods produce the same pollution, because of the wind the East neighborhood suffers a greater decline in air quality than the West.

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 Figure 1. The disamenity causes equilibrium rents paid by high-skilled workers to increase compared to the benchmark without pollution. Total amenities are affected in both neighborhoods, but the wind causes greater damage to the East. 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.

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. While that early industrial period set the industrial and urban landscape, 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 A3: 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 46.5% in 1950 (Crafts, 2014). Thereafter it declined, most rapidly in the 1980s when state-owned industries were privatized. By 2007, industry employed only made up 20.1% of total labor. 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 A3, these regulations had an immediate and marked impact on coal consumption.

The heavy reliance on coal between 1850 and 1950 generated an unprecedented concentration of sulphur dioxide, which scarred newly-industrialised cities and their surroundings. The negative impact of atmospheric pollution is captured in a classic 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, so 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).

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 and agricultural counties into the emerging industrial cities (see Shaw-Taylor and Wrigley, 2014). As shown in Williamson (1990) and

[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.

the Appendix Figure A2, the growth of cities peaked in the 1830s and then slowed down as the nineteenth century proceeded. By the end of the nineteenth century, the large cross-country migratory flows that marked the early Industrial Revolution had moderated significantly, and cities had reached a steady-state.[11]

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

4 Data

This section describes the construction of our measure of atmospheric pollution around 1880-1900 and neighborhood composition in 1817, 1881 and 1971-2011.

We will achieve this by addressing two challenges. 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 at the Lower Super Output Area definition of 2001.

4.1 Construction of the Air Pollution measure

Our strategy to generate a geo-referenced air pollution map for our 70 metropolitan areas covered by Ordnance Survey Maps can be summarized as follows: In a first step, we go through each geo-referenced map tile and mark each visible chimney with a unique identifier. We then use a recognition algorithm to locate each mark, and extract the associated identifier and its coordinates. In a second step, we run an augmented plume-diffusion model to 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, e.g., the Lower Super Output Area in 2001, and overlay all chimney-specific pollution imprints to generate a unique air pollution measure for each geographic unit. We describe these three stages in more detail below.

Identifying chimneys We rely on Ordnance Survey 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 from the 1840s to the

[11] For instance, Williamson (1990) shows that the portion of city growth due to migration declines over the nineteenth century. In a pioneering early study of the 1881 census, Ravenstein (1885) found that 75.2% of those recorded in England and Wales resided in the county of their birth.

1950s. The maps contain details on roads, railway, rivers, canals as well as on public amenities such as schools and parks. Most useful for our purposes, these maps also outlined the buildings of factories and, 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 as well as their various sizes prevent us from directly using a recognition algorithm (two examples of symbols are shown in Figure 2). Instead, we went through all map tiles and marked chimneys with a recognizable symbol X and a unique numeric identifier. This allowed us to retrieve the geolocation of each chimney. The maps also include information about the factory's sector. This allows us to assign to each chimney a specific sector and then use its coal intensity in pollution modelling.

An example of the chimney-identification is provided in Figure 3. On this map fragment, four different chimneys can be identified.[12] The symbol X is located in the center of a chimney and is used to geo-locate the chimney. The 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*. The red symbol X can then be identified by a recognition algorithm which, together with the projection provided by the Ordnance Survey, allows us to geolocate each chimney.

We restrict our analysis to 70 cities in England (see the online Appendix Figure A11) and their metropolitan areas. These cities constitute a quasi-exhaustive snapshot of industry and its associated pollution. These metropolitan areas comprise about 60% of the total population in 1801 and 1891 and 66% of the total population in 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. In particular, it models atmospheric dispersion under a large spectrum of meteorological conditions in addition to wind intensity and direction. For example, it allows us to refine the

[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.

pollution estimates in coastal areas and to incorporate 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 industrial chimneys during the Industrial Revolution were at a much lower altitude than modern chimneys, their 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 the slight changes related to climate fluctuations between the 19th century and today.[14] Second, *ADMS 5* requires complex terrain data and potential 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.[15] 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. In 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. We use conservative values for the emission source and assume 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 Figure 4 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). Even within a given day, wind conditions may be very unstable (i.e. not have a predominant direction) and the smoke may drift very little from its original source.

We report sensitivity analyses where we vary the two key parameters (exit velocity and chimney height) in the online Appendix. In Appendix Figure A9, we report 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. Our benchmark measure uses an average of these conditions over the past 10 years. In Appendix Figure A10, we display a synthetic atmospheric dispersion map for a point source

[14] Long-term and cyclical fluctuations in the Gulf Stream are not included in our model.

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

with flat terrain in North England under two scenarios: (a) The benchmark chimney height of 25 meters, and, (b) a taller chimney of 40 meters. While the shapes of both clouds are very similar, the second cloud affects the immediate neighborhoods much less and reaches further than the benchmark specification.

We finally 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.

Aggregation at a given geographical unit Atmospheric dispersion models are additive. Total Air Pollution measures are computed as the sum of each separate chimney (as well as household emissions). 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 (glass etc.). 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.[16]

Figure 5 displays the industrial sources of pollution for Manchester and Oldham (left panel) and the resulting aggregate Air Pollution (right panel). We can see that the pollution cloud tilts toward the East. We finally collapse our data at the level of 2001 Lower Super Output Areas to assign our pollution measure to persistent spatial units.

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. We would like to use address information but geo-referencing historical addresses is

[16] We will show specifications with an unweighted measure in the robustness checks.

not a straightforward exercise because street names often changed over time. 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. If we can confidently 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* the geo-located parish boundaries.

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

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 our 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).[18]

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.

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

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

We then repeat this algorithm with different cluster definitions, compare the resulting LSOA identifier under the different specifications, and select the most likely LSOA identifier. Sensitivity checks are also discussed in the Appendix.

Neighborhood composition in 1817, 1881 and 1971-2011 For 1817, we use "The Occupational Structure of England and Wales, c.1817-1881" (Shaw-Taylor and Wrigley, 2014) in which they use baptism records over 1813–20 to reconstruct a quasi-census for 1817. We form 834 parishes, that can be consistently linked with parishes in 1881. For 1881 we use the occupational census at that year. With these data, we observe counts of adult males in 539 occupations, and we distinguish broad social categories: Capitalists; managers; professionals; employees/clerks; skilled; semi-skilled; unskilled; farmers and special categories such as military.

In 1971, 1981, 1991, 2001 and 2011, we use census aggregate data at the smallest geographic level (i.e. the enumeration district or census output area in later years), and collapse all data at the 2001 Lower Super Output Area (LSOA) level to generate persistent geographic units between census waves.[19] The census data provide consistent measures of occupation, housing and country of origin for all these years.

One drawback of the Censuses 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 or parish income from average occupational wages and rentiers' income. However, such inference would require strong assumptions especially regarding the relative occupational labor income 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 share of low-skilled workers.[20]

For 1817 and 1881, we first collapse our over 500 occupational categories into 11 categories: Farmers, unskilled workers; semi-skilled workers; skilled manual workers; unemployed; disabled; soldiers; rentiers; gentlemen; managers; and, clerks. We then define low-skilled workers as the manual unskilled and semi-skilled workers, and the

[19] If enumeration districts (census output areas) related to more than one 2001 LSOA we would split them proportionately to their share of land in the respective LSOAs.

[20] We calculate the share of low-skilled workers as share of low skilled workers in an LSOA. We can additionally standardize this share by the share of low skilled workers in the city. Doing so does not change our results. In the follwing, we use the share of low skilled workers in an LSOA since it is easier to interpret.

job seekers (representing a very small fraction). 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 further refine our measure, we restrict our sample to individuals with the lowest possible measurement error, i.e., males between 25 and 55.[21] 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 urban centers and their immediate surroundings.

For 1971-2011, the occupational categories are already classified into 1-digit clusters: Managers; professionals; associate professionals; administration; skilled manual workers; care and leisure services; sales; processing/machinery; and, elementary. We replicate our classification in the main categories for males between 25 and 55 with two modifications. (i) 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 skilled manual workers are thus classified as low-skills, which brings about 62% of low-skills, 38% of high-skills in 2011. (ii) We drop the category "farmers" as it is almost non-existent among our modern, urban LSOAs.

In the robustness checks, we consider alternative indicators of neighborhood composition for the recent period. In particular, we use house prices as recorded by the Land Registry and explore the different components of deprivation with the English Indices of Deprivation (2010) sub-indices (Income; Employment; Health and Disability; Education, Skills and Training; Barriers to Housing and Services; Crime; Living Environment).[22]

Limiting the sample to neighborhoods within a buffer of 20 kilometer around our 70 cities, we obtain social composition measures for about 4,500 LSOAs and 834 consistent parishes between 1817 and 1881. Large cities such as Bristol, Liverpool or Manchester are covered by almost 100 LSOAs each. This sample is the most conservative. With fewer restrictions on the clustering process or the fuzzy matching, our sample size would increase at the expense of including LSOAs with higher measurement error.

4.3 Descriptive statistics

Table 1 provides summary statistics for the full sample and, for each variable, a decomposition of the variance within and between cities.

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

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

The clustering process described above classifies about 10 million individuals in $4,524$ LSOAs in 1881. As these LSOAs are the 2011 Census units, we can associate contemporary measures for all $4,524$ observations, and we only lose 5 observations when we create topographic controls.

In the first lines of Table 1, 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 A4 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 or Liverpool) have indices of pollution above the peaks recorded in contemporary Beijing ($40\mu g/m^3$).

We also provide in Table A11 an illustration of the within-city variation in air pollution. We compare our estimates in 10 neighborhoods to the deposits collected by the First Annual Report of the Sanitary Committee on the Work of the Air Pollution Advisory Board, 1915.[23] 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 correlate very strongly with the deposits collected in 1915.

5 Empirical strategy

5.1 Benchmark specification

To estimate the impact of pollution on neighborhood sorting within cities, we would ideally use a difference-in-difference specification and identify the sorting in response to pollution from differences between the pre-treatment period (1817) and the post-treatment one (1881).

However, our measures of segregation in 1817 are only computed at the parish level and a proper difference-in-difference specification would therefore require using

[23] 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.

the parish as the unit of observation throughout. Since cities like Bristol, Liverpool or Manchester are covered by about 90 LSOAs but only 10 to 20 parishes, we would lose a lot of information. Therefore, we leave the difference-in-difference specification with all variables collapsed at the parish level as a robustness check and use simple difference specification at the LSOA level in $t = 1881, \ldots, 2011$ where we control for parish-level characteristics in 1817 as our benchmark. Reassuringly, the two specifications provide very similar results.

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:

$$Y_{it} = \alpha + \beta P_i + \gamma \mathbf{X}_i + \nu Y_p + \delta_c + \varepsilon_{ict} \tag{7}$$

where Y_{it} are our measures of occupational structure, P_i is the normalized pollution as predicted by our chimney locations and the pollution dispersion model, $\mathbf{X}_i$ are geographic controls (elevation, distance to the town hall, longitude, etc.), Y_p is the occupational structure in 1817 at the parish level and δ_c are city Fixed-Effects.

A concern with specification (7) is that our treatment may not be fully exogenous conditional on controls $\mathbf{X}_i$ or even Y_p, because some unobserved amenities may explain both the presence of industries and the occupational structure in some neighborhoods. For instance, neighborhoods with low amenities may attract large polluting factories together with low-skilled migrants. In a robustness check, we will estimate a variation of (7) controlling for synthetic pollution generated by a "static" wind. The identification then only relies on the asymmetry between neighborhoods equidistant from factories, some of them being located in the East Side versus the West Side. This specification will account for direct effects associated with proximity to factories.

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

5.2 Controlling for non-random industry location

In order to clean our treatment of potential non-random industry location, we construct synthetic pollution imprints that draw on exogenous variation in pollution sources interacted with the exogenous atmospheric dispersion. In other words, we interact the exogenous variation underlying the choice of industry location with the exogenous wind flows carrying the resulting pollution to neighborhoods.

We suggest two different ways to obtain exogenous variation in industry location.

In a first specification, we exploit the fact that chimneys required large boilers and so a constant stream of water for cooling. As a result, all mills are located along rivers or canals. We exploit this technical necessity using canals and rivers observed in 1827, before the rise of coal as main energy source. As a result, they were not susceptible of being selectively placed upwind of poor neighborhoods because of resulting pollution. However, this variation also correlates with distance to canals which may itself affect the attractiveness of a neighborhood. We thus need to interact this variation with the exogenous atmospheric dispersion as predicted by the existing meteorological conditions, and control separately for distance to canals. We create the interaction of these two variations by locating synthetic chimneys uniformly along all existing canals and rivers in 1827, and we model atmospheric dispersion from these counterfactual pollution sources(see an example in The Appendix Figure A5).[24]

In a second specification, we create a pollution prediction cleaned from variation in pollution sources within cities. We locate synthetic chimneys uniformly within the 1890 city borders such that the resulting pollution prediction only relies on variations in the urban shape, topography and wind patterns.

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 segregation:

$$\begin{cases} P_i = b_0 + b_1 PP_i + \mathbf{c}\mathbf{X}_i + d_c + fY_p + e_{ict} \\ \\ Y_{it} = \beta_0 + \beta_1 \widehat{P}_i + \gamma \mathbf{X}_i + \delta_c + \nu Y_p + \varepsilon_{ict} \end{cases} \tag{8}$$

where Y_{it} are our measures of occupational structure in a particular Census wave ($t = 1881, \ldots, 2011$), PP_i is one of the two synthetic treatments predicted by chimneys located along the 1827 canals or a uniform distribution of chimneys, P_i is the pollution as predicted with actual chimney locations, $\mathbf{X}_i$ are geographic controls (elevation, distance to the town hall, distance to canals, geographic coordinates etc.), Y_p is the occupational structure in 1817 at the parish level and $\{\delta_c, d_c\}_c$ are city Fixed-Effects.

[24]The online Appendix Figure A6 describes our approach. In panel (a), we see the cities of Manchester (left) and Oldham (right) with the associated 1827 canals and rivers. Panel (b) displays the counterfactual chimney locations and panel (c) the resulting air pollution. Finally, panel (d) shows the actual pollution.

6 Historical pollution and neighborhood sorting

In this section, we describe our main empirical results and document new stylized facts about air pollution and neighborhood sorting.

We first document the negative correlation between historical atmospheric 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 a same city. While we control for important neighborhood characteristics in our benchmark specification (longitude and latitude, distance to main amenities including canals 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.

6.1 Benchmark results

We start with our baseline specification to capture the relationship between air pollution and neighborhood sorting.

In Table 2, we report the estimates for specification (7) 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 highly 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 Tables 2, we add city fixed-effects to control for the variations between cities in atmospheric pollution and occupations. In the third column, we condition our estimates on elevation and distance to rivers or canals to control for potential confounders between neighborhood sorting and pollution. In the fourth column, we add the parish-level shares of low-skilled, high skills and farmers in 1817

to clean for potential unobserved LSOA fixed characteristics. In the fifth column, we control for the position of LSOAs within the city. In the sixth column, we add latitude and longitude of the LSOA centroids to control for potential Western or Southern preferences in locations. As apparent from Table 2, none of these controls affect our estimates.[25]

Figure 6 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 of the (standardized) shares of low-skilled workers in 1881 on the same set of controls as in column 4 of Table 2. On the x-axis, we plot the 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 and very high 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 section, we address this issue and present further other robustness checks.

6.2 Robustness checks

We now 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 a non-random and strategic location of industries within cities.

First, we show in Table 3 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 both statistically and economically. This placebo check is reassuring since it suggests that potentially unobserved, pre-existing neighborhood characteristics are not driving our results.

Second, we run difference-in-difference specifications either at the LSOA-level (attributing the average parish-level share of low-skills in 1817 to LSOAs) or at the parish-level. As shown in the online Appendix Table A2, the difference-in-difference estimates at the LSOA-level are very similar to the simple difference estimates of Table 2. The estimates for the difference-in-difference specification at the parish-level (see Appendix Table A3) are slightly larger, but less precisely estimated. This approach, along with the balancing test, reduces 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 the future city development. We tackle this issue in the following tests.

[25]The first column of Appendix Table A1 reports coefficients on the covariates.

Third, we generate three counterfactual pollution imprints from actual industry locations but alternative air pollution profiles.[26] 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 a mirror wind profile (rotation of 180° around the source) going from the North-West to the South-East (*Mirror* pollution). Note that the large number of chimneys across the city implies that a measure like the distance to the closest chimney carries no additional information.[27] 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 captures the proximity to a dense cluster of factories. As shown in Table 4—columns 1 and 2, none of these counterfactual atmospheric pollution measures affect our estimates. If anything, our estimates are larger illustrating that the mere proximity to chimneys may induce some positive neighborhood sorting. In column 3, we control for domestic pollution as predicted by the location of private residential buildings across the city, which does not affect our estimates. We also add contemporary pollution in 2011 as constructed by the Department for Environment, Food and Rural Affairs (DEFRA). Due to deindustrialisation and new sources of emissions, the past and contemporary pollution clouds are very different and the correlation between the two atmospheric pollution measures is low.

Fourth, 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 2 – 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 (about 1270 in our sample) and in column 3 to Medium Lower Super Output Areas (the geographic unit just above LSOA in the census, about 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.[28] In Panel B, we report standard errors clustered at three different levels, electoral ward, MSOA and city. Standard errors increase by about 40% between the least and most conservative

[26]In an unreported robustness check, we verify that a measure of air pollution based on a uniform weight for each chimney as opposed to weighting depending on the sector of the factory gives the same results. Our estimates are quantitatively and qualitatively similar with these more conservative chimney weights.

[27]We can, however, control for the number of chimneys in each LSOA. As for the two counterfactual pollution patterns, this control does not change the baseline estimates at all (unreported specification).

[28]There are 4 Lower Super Output Areas on average within each MSOA in our sample.

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.

Finally, we present in Table 5 the results of our 2-stage strategy, that uses 1827 canals as a source of exogenous variation for chimney location (columns 1 and 2) and a random allocation of chimneys within 1890 city boundaries (columns 3 and 4). The first stage is very strong in both cases given that (i) many industries tend to locate around canals, and (ii) wind patterns interacted with topography play a very big role in the atmospheric dispersion of pollutants. 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 10 percentage points in the 2-stage strategy with 1827 canals, and 6.5 percentage points in the 2-stage strategy with uniformly-distributed factories. These results may indicate that the non-random component of industry location is associated with positive neighborhood sorting in the late nineteenth century.

7 Dynamics of persistence after the Clean Air Acts (1971-2011)

This section focuses on the relationship between historical atmospheric pollution and neighborhood income 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 our previous analysis of neighborhood segregation in 1881 to recent waves in 1971, 1981, 1991, 2001 and 2011. Table 6 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. 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 that is often evoked in popular culture.

We present in the Appendix two robustness checks with (i) alternative measures of deprivation and (ii) house prices. In Appendix Table A5, we use the composite deprivation subindexes (income, employment, education, health, barriers to housing and crime). 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 almost all scores. The only subindex 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 the Land Registry of all transactions in England and Wales between 2000 and 2011 and create the (logarithm of the) average transaction prices and number of transactions between 2000 and 2011 as well as the average shares of detached, semi-detached, terraced houses, and flats/maisonnettes. The first (resp. last) two 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 types. 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% (9% of a standard deviation). Looking at transactions, we find that an additional standard deviation in past pollution is associated with an 8% (14% of a standard deviation) decrease in the number of transactions and 15% (28% of a standard deviation) when controlling for house types. We illustrate the relationship between house prices and pollution (as estimated in Table A6 – column 1) in the Appendix Figure A7.

The persistence of the relationship between historical pollution and neighborhood segregation cannot be mechanically attributed to the collapse of industries in the former *cottonopolis*. Indeed, our estimates are identified within cities, and unreported specifications where we rerun the sensitivity analysis in the previous section for the contemporary measures of social composition are robust to controlling for distance to former factories, a larger set of fixed effects or the 2-stage specification.[29]

In order to visualize potential non-linearities in the persistence of neighborhood sorting, we display in Figure 7 the relation between the shares of low-skilled workers in 1881, 1971, 1981, 1991, 2001 and 2011 and the temporary disamenity. As apparent, there are no signs of returns to the mean for extreme values of pollution, i.e.,

[29]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.

one standard deviation above or below average within-city pollution. The gap with average neighborhood composition remains constant at the extremes. In contrast, we can observe a reversion to the mean for intermediate values of within-city pollution. This pattern would be consistent with the existence of tipping points leading to a higher persistence in neighborhoods with the most extreme pollution values.

The remainder of this section will build on a quantitative model to explain the non-linear persistence between past disamenity and neighborhood composition, and we will use the model to better understand the dynamics of segregation across cities (exposed to different intensities of the initial disamenity, pollution).

7.2 A quantitative, dynamic model of sorting

In Section 2, we laid the foundations of the quantitative model in a static framework of neighborhood sorting. Here, we extend the model to a dynamic framework where the persistence of sorting is rationalized by an endogenous amenity.[30]

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.[31]

The purpose of the quantitative model is not to posit micro-foundations for persistent neighborhood effects but rather 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, $\bar{\theta}(j,t)$[32]. Motivated by the possibility of tail effects, we model a component, $b(j,t)$, that reduces the attractiveness of

[30] 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.

[31] Durlauf (2004) and Rosenthal and Ross (2015) are excellent overviews of the range of neighborhood effects that may play a role. If income levels differ, neighborhoods could accumulate amenity differences that then have long-run consequences. 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).

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

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,

$$d(j,t) = (1-\delta)d(j,t-1) + e(j,t) - b(j,t), \tag{9}$$

with an initial endogenous amenity that is constant across all neighborhoods, $d(j,1) = d$, and where the continuous effect is,

$$e(j,t) = \phi_1^e \left[\bar{\theta}(j,t-1) - 1\right]^{\phi_2^e}, \tag{10}$$

if $\bar{\theta}(j,t-1) > 1$ and 0 otherwise. The tail effect captures a discontinuity in neighborhood income,

$$b(j,t) = \phi_1^b \left[1 - \bar{\theta}(j,t-1)\right]^{\phi_2^b}, \tag{11}$$

if $S^l(j,t-1) \geq \bar{S}$ and 0 otherwise. 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 the accumulation of factors that affect neighborhood choices, or if the pollution caused neighborhoods to cross a threshold in the share of low-skilled workers, 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 $d(j,t)$,[33]

$$S^l(j,t_2) - S^l(j,t_1) = \underbrace{-\alpha p(j)}_{reversion} + \underbrace{\text{sign}\{p(j)\} \cdot d(j,t_2)/2}_{persistence} \tag{12}$$

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 (12) is linearly decreasing in $p(j)$. In this case, the initial pollution causes sorting but there is later full reversion to the mean. If instead $\phi_1^e, \phi_2^e > 0$, the continuous neighborhood effect acts to solidify the initial sorting. If there was less than average historic pollution in a neighborhood, this effect dampens 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. In Table 7, we report the model parameters that are first selected to match the data. We rely on Williamson (1980) for data on income inequality in nineteenth century England: We set the ratio of high income to low income at two.[34] Finally, we use the correlation between within-city residuals in low-skills and atmospheric pollution in 1871 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.[35] In Table 8, we report the parameter estimates that minimize the root mean squared error between the model prediction and the data for the change in low-skill share between 1881 and 1971. 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

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

[34] 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.

[35] 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 8).

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 operates once a neighborhood is 26 percentage points higher in low-skill share than the city average. Finding a value of the exponent ϕ_2^b that is greater than one suggests that the tail costs are convex. In general, the tail effect is stronger than the reversion process and there is no return to the mean – once a neighborhood has suffered from enough pollution to cross this threshold, its long-run outcome in terms of skill-share is worse 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",

$$\text{spread}(t) = E\left[S^l(j,t)\middle|p(j) > 0\right] - E\left[S^l(j,t)\middle|p(j) < 0\right] \tag{13}$$

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,

$$\rho_{t_2,t_1} = \frac{\sum_j \left(S^l(j,t_2) - \gamma\right)\left(S^l(j,t_1) - \gamma\right)}{\sum_j \left(S^l(j,t_1) - \gamma\right)^2} \tag{14}$$

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

The first two columns of Table 9 report results on these statistics in the model and the data. The model performs well in matching the 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.

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 Mrs Thatcher's reform of social housing. We extend the model to consider this.

Liberalization of social housing Before 1979, social housing was distributed relatively uniformly across neighborhoods (see the online Appendix Figure A8).[36] In 1979, Mrs 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[37] 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 higher (lower) 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 9 reports model output with social housing liberalization ('SH-L' in the Table) against the baseline model and the data. The model fit for 1971-2011, either captured by the 2011 spread or the correlation between 1971 and 2011, substantially improves: the liberalization of social housing caused greater persistence in the distribution of deprivation because it removed a random component (the location of social housing) which was bringing neighborhoods closer to the city average. With the liberalization, these randomly-allocated low-skilled workers would relocate in deprived neighborhoods.

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 grows steadily over the period with much

[36] The United Kingdom initiated a programme 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 main supply of housing services for the working class, and it was typically managed by local councils. About 30% of our sample of urban households were living in council houses in 1971.

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

fewer owners in the areas that were formerly affected by coal pollution.[38]

While the original intent of Mrs Thatcher's policy was to reduce inequality by providing a route for working class households to step on the housing ladder, its consequence appears to have been to lengthen the shadow of the Industrial Revolution and set back the slow decay of neighborhood sorting. With our estimates, about 20% of the remaining gradient between Western and Eastern neighborhoods could 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 10, 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 programmes 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.[39] As can be seen in columns 4 and 5 of Table 10, a 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 the last theoretical prediction. Cities in our sample have similar shares of low-skilled workers but they differ widely in exposure to pollution. We define

[38] 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).

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

$share_{polluted}$ 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 60%).

We report in Table 11 the benchmark 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 16% in 1881 and 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 return to the mean. The standardized effects are around 13% in 1881 and 25% in 1971, and then range between 23 and 29% in following waves. The two sets of cities respectively stand 25% lower and 30% higher than the average pollution exposure, such that we can compare the persistence in the two sets of cities to the columns 4 and 5 of Table 10. In Table 10, persistence in lowly-exposed cities is 2/3 lower than in highly-exposed cities while it is half as low in Table 11.

In order to illustrate the dynamics of persistence in the two sets of cities, we represent graphically the relationship between pollution and indices of deprivation across waves (see Figure). In Panel (a) (resp. (b)), we display the distribution of pollution across cities with a high (resp. low) share of polluted areas. In Panel (c) (resp. (d)), we display the respective evolution of the relationship between deprivation and past pollution. There is a clear difference between Panel (c) and (d). In 1881 or 1971, the relationship between atmospheric pollution and deprivation is quite comparable in both samples of cities. In recent waves, however, the relationship is stronger in cities where the distributions of pollution and 1881 deprivation coincide, and this amplification seems to be driven by the extremes. In contrast, the relationship is weaker in the other cities (panel (d)), which seems to be essentially due to a return to the mean for intermediate values of pollution. While this analysis is not causal – the two sets of cities may differ across different dimensions, these findings are consistent with the theoretical prediction: cities with a clean initial sorting are much more rigid.

Finally, a shortcoming of our analysis is that we cannot identify the different mechanism 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 a brief exploration of two candidate mechanisms, i.e., durable public amenities and housing stock. We do not find a strong support for any of the two channels.

8 Conclusion

This paper provides an explanation for why the East Sides of formerly-industrial cities in the Western hemisphere tend to be poorer than the West Sides. *Westerlies* blew from West to East and, with rising coal use in the heyday of the industrialization, pollution became a major disamenity in the East Side. This unequal distribution of pollution exposure induced a sorting process which left the middle and upper class in the relatively less polluted West and the working class in the East. Our empirical analysis relies on very precise pollution estimates, and identifies neighborhood sorting at a very 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 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 sustaining the inertia between 1881 and 1971. Our estimates imply large non-linear effects with tipping-like behaviors, 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 depends on how long these areas have been segregated. 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 cost of an uneven pollution exposure across space: pollution may induce large spatial inequalities that survive de-industrialisation.

References

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

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

Barreca, Alan, Karen Clay, and Joel Tarr, "Coal, Smoke, and Death: Bituminous Coal and American Home Heating," Working Paper 19881, National Bureau of Economic Research 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.

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," Working Paper 22155, National Bureau of Economic Research 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., "Economic growth during the long twentieth century. Chapter 2 in Floud, R., Humphries, J. and Johnson, P. (eds.)," *The Cambridge Economic History of Modern Britain: Volume 2, 1870 to the Present*, 2014. Cambridge University Press.

_ **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.

Davis, Donald R. and I. Dingel Jonathan, "The Comparative Advantage of Cities," *NBER Working Paper Series*, 2014, (20602).

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken, "What Do We Learn from the Weather? The New Climate-Economy Literature," *Journal of Economic Literature*, 2014, *52* (3), 740–98.

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.

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

_ , "Chapter 50 - Neighborhood Effects," in J. Vernon Henderson and J.F. Thisse, eds., *Handbook of Regional and Urban Economics*, Vol. 4 of *Handbook of Regional and Urban Economics*, 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," University of Ottawa Working Paper No 1514E 2015.

Hornbeck, Richard, "The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe," *American Economic Review*, 2012, *102* (4), 1477–1507.

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.

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.

Maw, Peter, Terry Wyke, and Alan Kidd, "Canals, rivers, and the industrial city: Manchester's industrial waterfront, 1790–18501," *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.

Neumark, David and Helen Simpson, "Chapter 18 - Place-Based Policies," 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. 1197 – 1287.

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," mimeo. 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. (eds.)," *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.

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. Paper 1 at www.hpss.geog.cam.ac.uk/research/projects/occupations/britain19c/papers/.

A Figures and tables

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

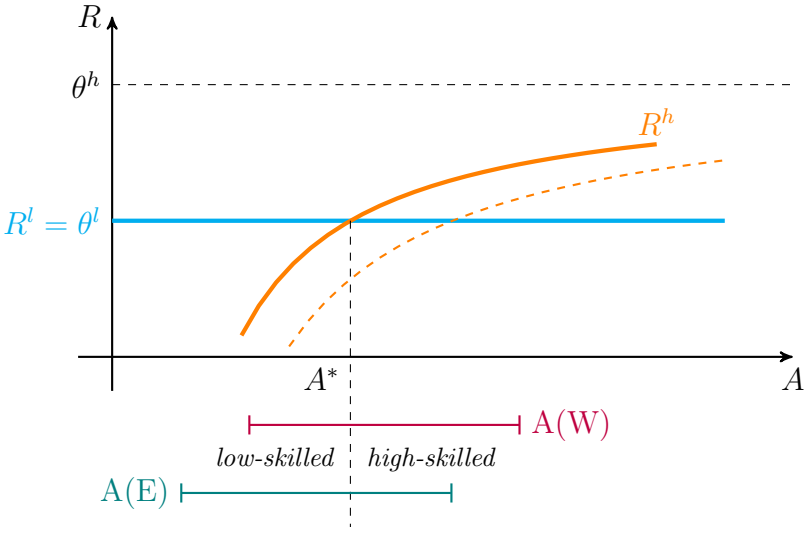
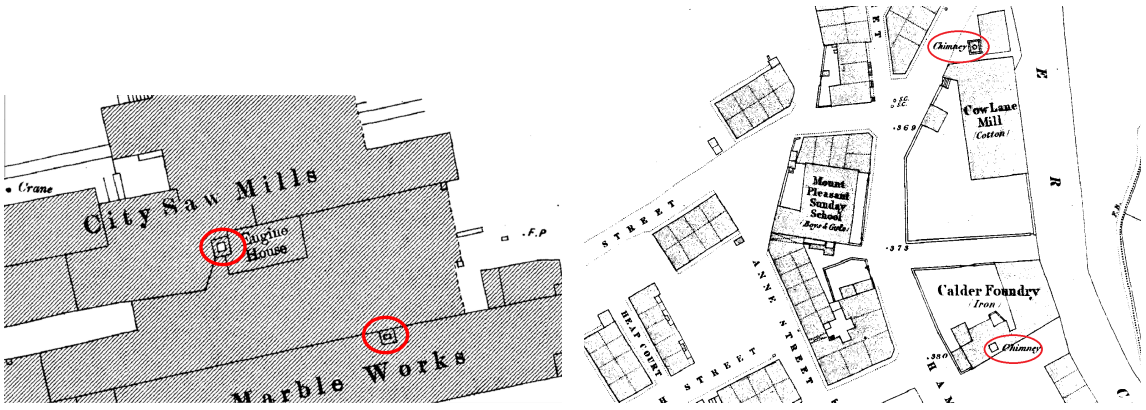


Figure 2. Town maps – chimney symbols.



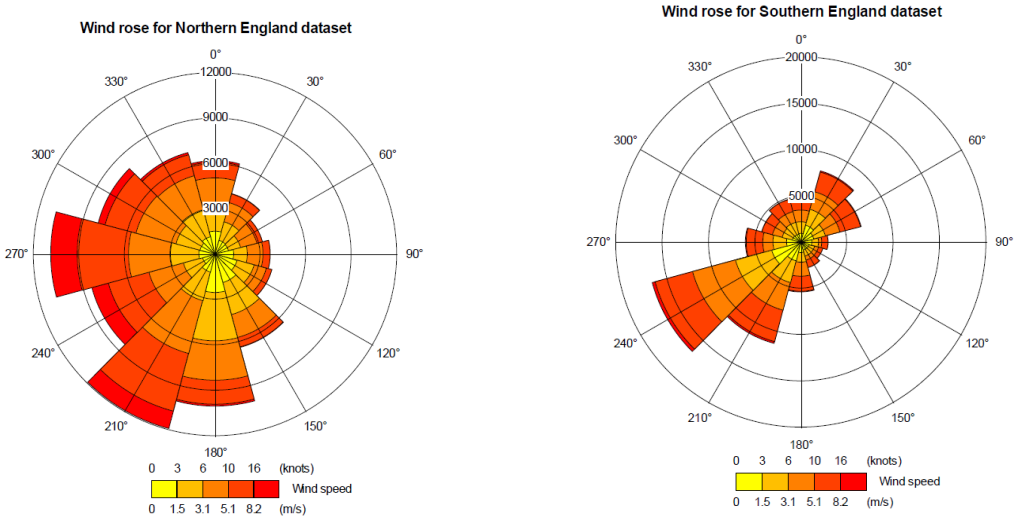
(a) Example 1. (b) Example 2.

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

Figure 3. Town maps – marking and identifying chimneys.

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

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

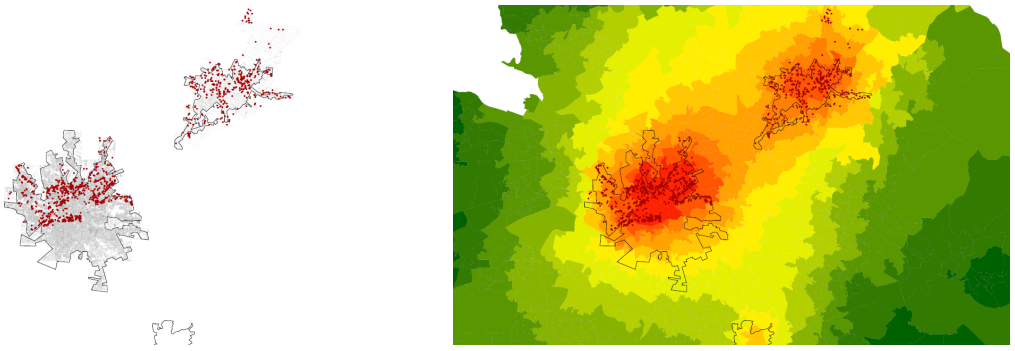


(a) North England.

(b) South England.

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

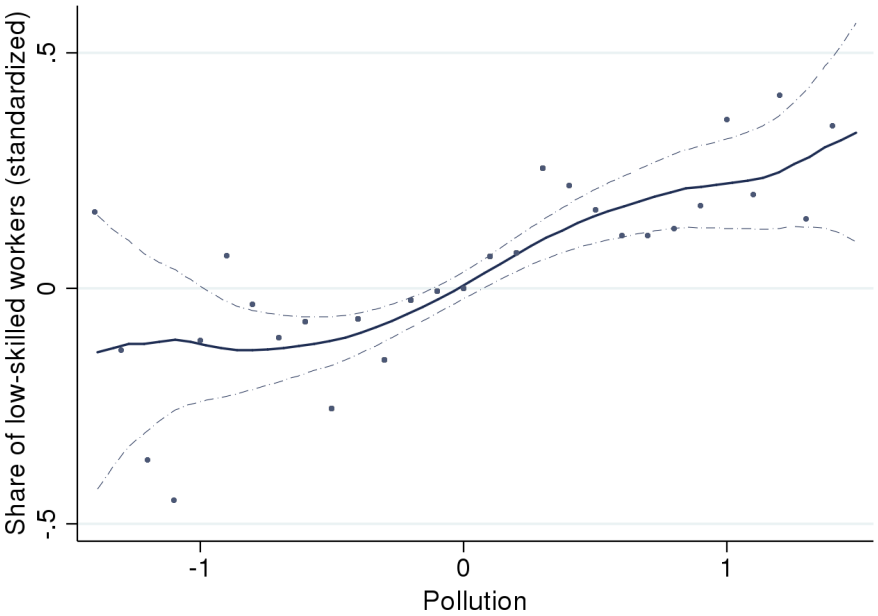
Figure 5. Aggregating pollution sources.



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

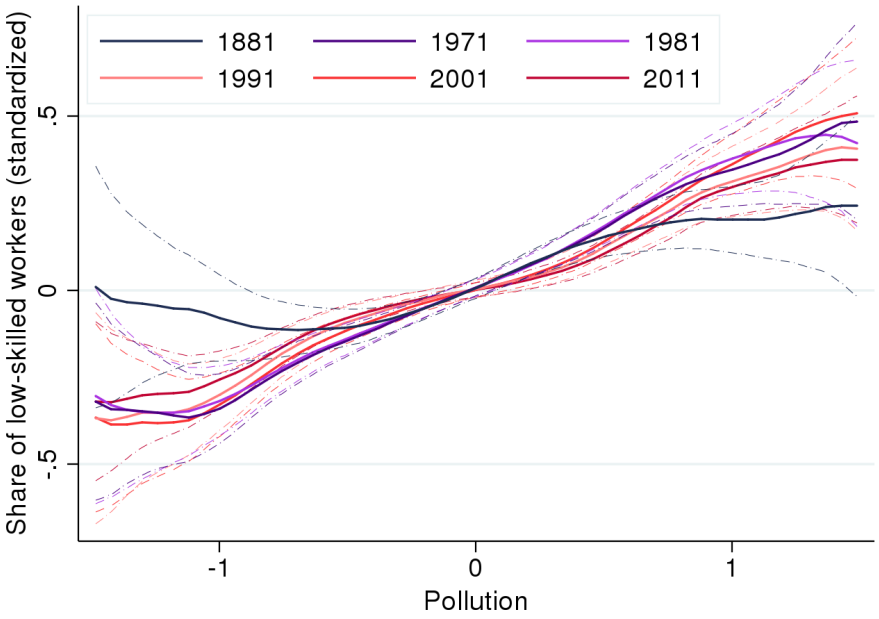
Sources: Authors' calculations using Ordnance Survey Maps - 25 inch to the mile, 1842-1952 and the ADMS 5 Air Pollution Model. Chimneys are indicated with red dots.

Figure 6. Share of low-skilled workers (y-axis) and pollution (x-axis) across neighborhoods in 1881.



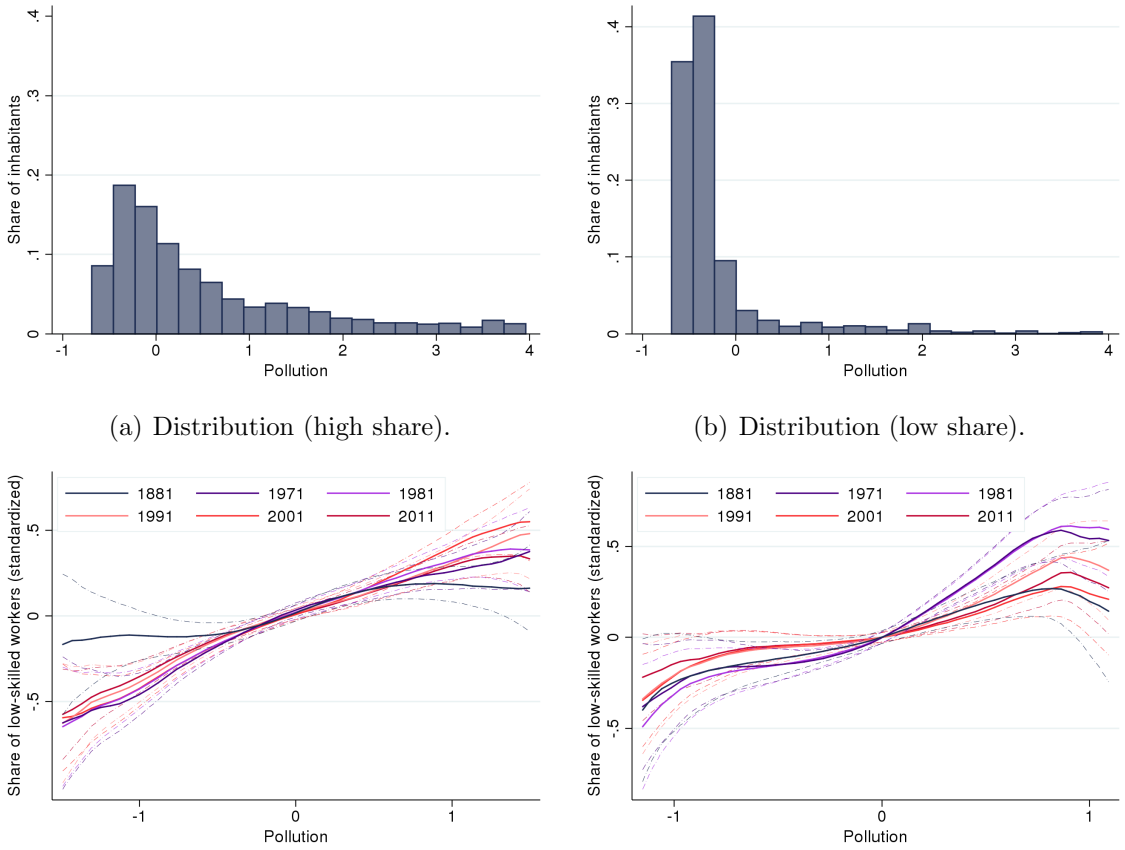
Notes: This Figure represents the relationship between the (standardized) shares of low-skilled workers in 1881 and our (standardized) measure of past pollution. We consider the residuals of all measures once cleaned by city Fixed-Effects, geographic and topographic controls. For the sake of exposure, 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 all observations.

Figure 7. 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.

Figure 8. 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.

Table 1. 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	.583	.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
		Geographic controls			
Distance town hall (m)	4,524	4823	5334	4754	1487
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 canals (m)	4,524	5723	14380	17899	1391

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

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

Share of low-skilled	(1)	(2)	(3)	(4)	(5)	(6)
Pollution	.0417	.0421	.0385	.0350	.0342	.0327
	(.0070)	(.0068)	(.0066)	(.0063)	(.0064)	(.0065)
	[.1686]	[.1700]	[.1557]	[.1415]	[.1381]	[.1321]
Observations	4,524	4,524	4,519	4,519	4,519	4,519
Fixed effects (city)	No	Yes	Yes	Yes	Yes	Yes
Controls (topography)	No	No	Yes	Yes	Yes	Yes
Controls (1817)	No	No	No	Yes	Yes	Yes
Controls (geography)	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 canals as of 1817. The set of 1817 controls include the parish-level shares of farmers, managers and blue-collar workers. 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. The set of geographic controls include distance to the city hall, share of LSOA within the city borders in 1890 and the area for the LSOA.

Table 3. Pollution and shares of low-skilled workers in 1817 – placebo check.

Share of low-skilled	(1)	(2)	(3)	(4)	(5)
Pollution	.0000	-.0048	.0123	.0052	.0062
	(.0125)	(.0196)	(.0235)	(.0245)	(.0241)
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

Robust standard errors are reported between parentheses. 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 1890 and the parish area. The set of topographic controls include the average, maximum and minimum elevations for the parish and the distance to canals as of 1817. 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 1817 is .78.

Table 4. 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	.0372	.0322	.0336	.0339
	(.0086)	(.0108)	(.0068)	(.0068)
	[.1505]	[.1302]	[.1356]	[.1370]
Mirror Pollution	-.0030			
	(.0062)			
	[-.0123]			
Static Pollution		.0021		
		(.0088)		
		[.0085]		
Domestic Pollution			.0087	
			(.0154)	
			[.0330]	
Current Pollution				.0106
				(.0056)
				[.0431]
Observations	4,519	4,519	4,519	4,519
Fixed effects (city)	Yes	Yes	Yes	Yes
Controls (geography)	Yes	Yes	Yes	Yes
Controls (topography)	Yes	Yes	Yes	Yes
Controls (coordinates)	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 geographic controls include distance to the city hall, share of LSOA within the city borders in 1890 and the area for the LSOA. The set of topographic controls include the average, maximum and minimum elevations for the LSOA and the distance to canals as of 1817.

Table 5. Pollution and shares of low-skilled workers in 1881 – 2-stage specification.

First stage	Pollution			
	(1)	(2)	(3)	(4)
Synthetic pollution (canals)	.3010	.2999		
	(.0093)	(.0094)		
Synthetic pollution (uniform)			.2497	.2478
			(.0090)	(.0090)
Observations	4,084	4,084	4,519	4,519
Fixed effects (city)	Yes	Yes	Yes	Yes
Controls (topography)	Yes	Yes	Yes	Yes
Controls (1817)	Yes	Yes	Yes	Yes
Controls (lat./lon.)	No	Yes	No	Yes
Second stage	Share of low-skilled workers (1881)			
	(1)	(2)	(3)	(4)
(Uniform) Pollution	.1017	.0995	.0683	.0644
	(.0169)	(.0172)	(.0190)	(.0194)
	[.4107]	[.4016]	[.2760]	[.2599]
Observations	4,084	4,084	4,519	4,519
Fixed effects (city)	Yes	Yes	Yes	Yes
Controls (topography)	Yes	Yes	Yes	Yes
Controls (1817)	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 topographic controls include the average, maximum and minimum elevations for the LSOA and the distance to canals as of 1817. The variable *Synthetic pollution (canals)* is only constructed for cities with some waterways in 1827.

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

Share of low-skilled workers	1971	1981	1991	2001	2011
Pollution	.0309	.0321	.0349	.0368	.0311
	(.0044)	(.0064)	(.0070)	(.0068)	(.0064)
	[.2397]	[.2239]	[.1841]	[.2278]	[.1776]
Observations	4,517	4,519	4,519	4,519	4,519
Fixed effects (city)	Yes	Yes	Yes	Yes	Yes
Controls (topography)	Yes	Yes	Yes	Yes	Yes
Controls (1817)	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 topographic controls include the average, maximum and minimum elevations for the LSOA and the distance to canals as of 1817. 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).

Table 7. 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 1871.

Table 8. Estimated parameters (see Section 7).

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

Note: The initial grid search is over the following ranges: $\phi_1^e = [0, 0.3]$; $\phi_2^e = [0, 1.5]$; $\phi_1^b = [0, 0.3]$; $\phi_2^b = [0, 2.5]$; $\bar{S} = [0.50, 0.90]$; $\delta = [0, 0.15]$.

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

	Data	Baseline	SH-L
Spread in 1971	0.0550	0.0555	0.0555
Spread in 2011	0.0278	0.0235	0.0281
Correlation $\rho_{2011,1971}$	0.4337	0.4010	0.4331

Note: SH-L is the baseline model augmented by the social housing liberalization of the Thatcher government in 1979 (See Section 7).

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

		Social housing		Pollution	
	Baseline	SH-40	SH-45	-25%	$+25\%$
Spread in 1971	0.0555	0.0555	0.0555	0.0421	0.0737
Spread in 2011	0.0235	0.0229	0.0218	0.0183	0.0358
Correlation $\rho_{2011,1971}$	0.4010	0.3698	0.3167	0.3885	0.5944

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

Table 11. 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 low 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
Controls (topography)	Yes	Yes	Yes	Yes	Yes	Yes
Controls (1817)	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
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. 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 canals as of 1817. 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).

Online Appendix

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,

$$\max_{j \in W,E} \left\{ \min_{\ell \in \Omega(j)} A(j,\ell) \right\} < \min_{j \in W,E} \left\{ \max_{\ell \in \Omega(j)} A(j,\ell) \right\}. \tag{15}$$

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,[40]

$$F(A) = \sum_{j \in \{W,E\}} A - \min_{\ell \in \Omega(j)} A(j,\ell), \tag{16}$$

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

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,

$$F(A) = 2(A - d + \rho). \tag{17}$$

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

[40] 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,

$$S^l(W) = \gamma - \eta\rho, \tag{18}$$

$$S^l(E) = \gamma + \eta\rho. \tag{19}$$

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 (10). The pollution, and consequent general amenities, may then cause $b(j,t)$ to accumulate by (11). 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,

$$A^* = \gamma + d(t)/2. \tag{20}$$

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

$$S^l(W,t) = \gamma - d(t)/2, \tag{21}$$

$$S^l(E,t) = \gamma + d(t)/2. \tag{22}$$

Since $d(j,t)$ amenities are persistent, $d(t) > 0$ permanently unless depreciation exists. Equations (21)-(22) then show that sorting persists even after $t = t_c$. □

B Geo-locating individuals in census data - detailed description

We describe here the Census structure, our 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.

Let $n : i \mapsto n(i)$ denote the block/neighborhood for individual i in a parish p. For each entry, we observe a book and a folio number summarized by $f : i \mapsto f(i)$, and we also observe the entry "line". We assume a monotonicity property since enumerators recorded sequentially: If two entries are in the same block, all entries appearing between these entries also belong to the same block.

This 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 books/pages $\{f_j\}_j$, 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 geographic cluster is n and one observed counterfactual is f. In what follows, we will describe our strategy as if folios were a perfect identifier for geographic proximity and we will discuss sensitivity analyses in a separate subsection.

Fuzzy matching of addresses We clean 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 separating the road denomination from the attributed name.

We reduce the probability of attributing a census address to the wrong georeference by limiting the pool of potential matches (the contemporary geo-located addresses) to those which are located in the registered parish of census observations. Any Census entry can thus only be matched to a contemporary address in the same parish.

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).

There is not a huge covariation among census entries in unmatched addresses, which indicates that most of the matching error comes from idiosyncratic sources. However, there remains some covariation 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 higher score than 0.90 and consider the others as being unmatched. We describe in the following section how we clean for potential errors in the already-matched household addresses and geolocate all remaining households using the subsample of households with a given geo-reference.

Recognizing clusters and inference We need to infer the geo-location of all households from (i) the geo-location of a subsample of them (with potential measurement error) and (ii) a proxy for geographic proximity (households with the same identifier). We first restrict our sample to the sample of well-matched households. We then apply the following algorithm to detect clusters.

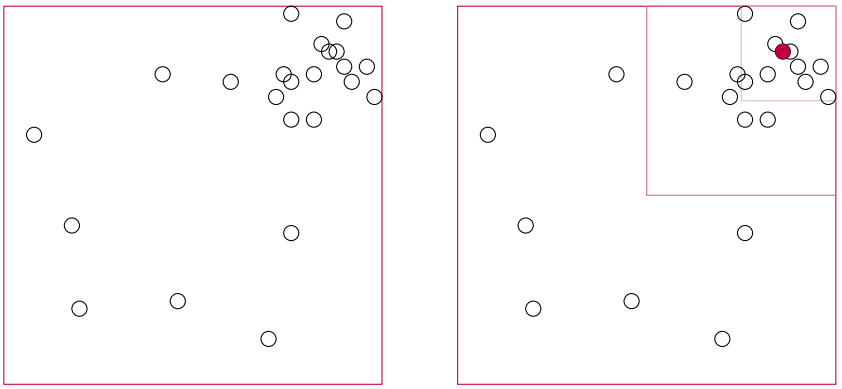
1. We locate all geo-referenced households within a parish.
2. We then divide a parish into 4 regions, depending on the position relative to the average between the maximum and minimum latitudes and the average between the maximum and minimum longitudes in the sample.
3. We select the region with the largest number of observations, and temporarily drop the other observations.
4. We go back to point 2 with the new selected region and new selected observations, and we iterate.
5. ...
6. After a given number of iterations, we stop, generate the average latitude and longitude among the remaining households and we attribute them to *all* households of the Census with the same identifier including thus the already-matched households.

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 situation, 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 use these newly-identified blocks, overlay them with our base geographic units (LSOAs) and attribute a unique LSOA identifier to all households in the 1881 Census.

Figure A1. Finding clusters among geo-referenced households with the same identifier.



Sensitivity analysis The previous methodology relies on two approximations.

First, folios are assumed to reflect underlying geographic identifiers. However, there is a tension between aggregating folios together (and have more households per identifier thereby being more likely to detect a cluster) or keeping them separate. Indeed, 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 practice, we repeat our algorithm by aggregating folios at different levels (grouping 2, 5 or 10 folios together) 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 is silent about the nature of the neighborhood effects that may operate, e.g., peer effects or durable amenities such as school quality or other public services.

In this section, we provide some empirical evidence on two potential mechanisms. First, the apparently persistent impact of past pollution may be related to the accumulation of durable public amenities, constructed during the industrial period in richer or less polluted neighborhoods. The proximity to schools, hospitals, parks

or public administration may anchor segregation beyond the initial (dis)amenity at the origin of neighborhood sorting. Second, the persistent impact of past pollution may be due to the housing stock and its persistence over time.

Public amenities We collect two sets of durable amenities around 1881 and 2011 in order to assess the role of potential landmarks in anchoring neighborhood composition. For instance, cities may locate public goods with positive local externalities in richer neighborhoods. For each LSOA, we compute the number of such public services per inhabitants (schools and universities, theaters, museums and libraries, parks, churches and hospitals) using the sets of 1880-1900 Ordnances Survey maps. In parallel, we use a directory of public, transport, health, education and leisure services in order to construct amenities in 2011 (in addition to the previous list, we include: local and national government buildings, courts and police stations, bus or train stations, botanical gardens and zoos).

As shown in Table A8, controlling for amenities in 1881 or 2011 does not affect the relationship between deprivation and pollution. While this observation may be expected with 1881 amenities (with some inertia in the location of public services), it is less evident in 2011. In order to better understand the joint distribution of past pollution, deprivation and amenities, we investigate the relationship between past pollution and these amenities in Table A10. There is no systematic correlation in 1881. In 2011, formerly polluted neighborhoods have more parks, recreational areas and transport facilities, and less hospitals, botanical gardens or conference centers. However, the estimates are quite small in magnitude. There may be two forces playing between 1881 and 2011: demand for high-quality amenities in good neighborhoods (e.g., hospitals) and low land prices in poor neighbourhoods (parks and recreation areas).[41]

In conclusion, amenities do not seem to play any role.

Housing age We collect house ages at the 2001 LSOA resolution from the Consumer Data Research Centre (CDRC). These data provide an exhaustive picture of building age in England and Wales but constitutes unfortunately a snapshot in 2011. In other words, this measure is relevant in 2011 but we cannot trace back the age of dwellings that have been recently destroyed and replaced in order to provide a snapshot in earlier periods.

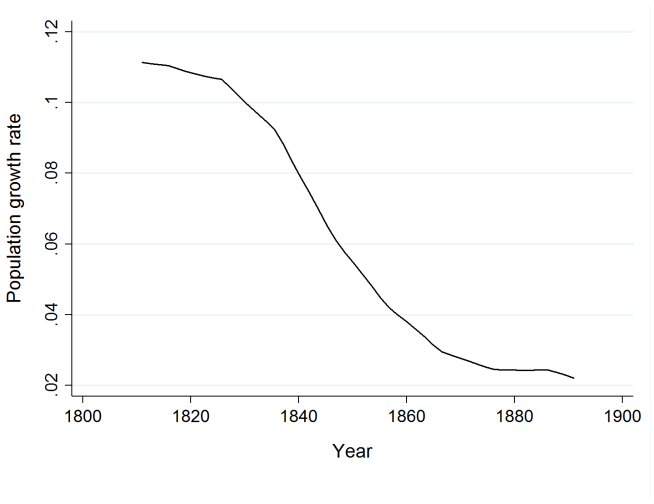
As shown in column 1 of Table A9, controlling for the building age in LSOAs does not affect the 2011 estimates reported in Table 2. In columns 2 and 3, we

[41] The presence of parks and recreation areas in formerly polluted neighborhoods may be the due to former industrial sites being destroyed in the second half of the twentieth century.

restrict the sample to LSOAs with a majority of new dwellings. While this selection may be spurious, it helps to understand if the pollution gradient can also be seen across neighborhoods with a recent housing stock. In column 2 (resp. 3), we restrict the analysis to LSOAs with a majority of dwellings constructed after 1940 (resp. after 1970). In both cases, the pollution gradient decreases by a notch but the standardized effect remains between .17 and .20.

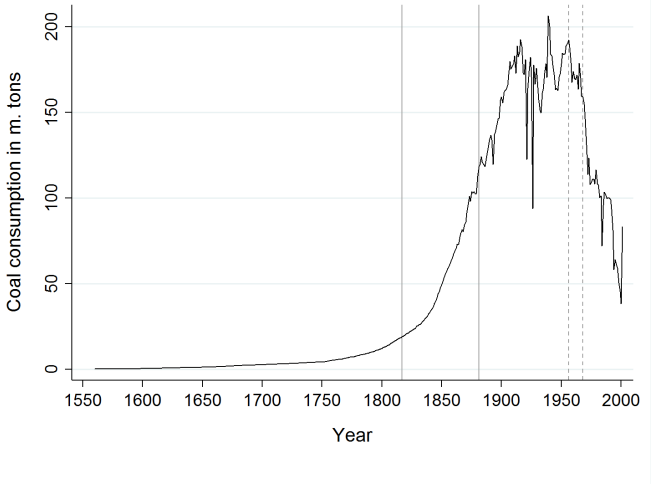
D Additional figures

Figure A2. Population Growth Rates in cities, 1801-1891.



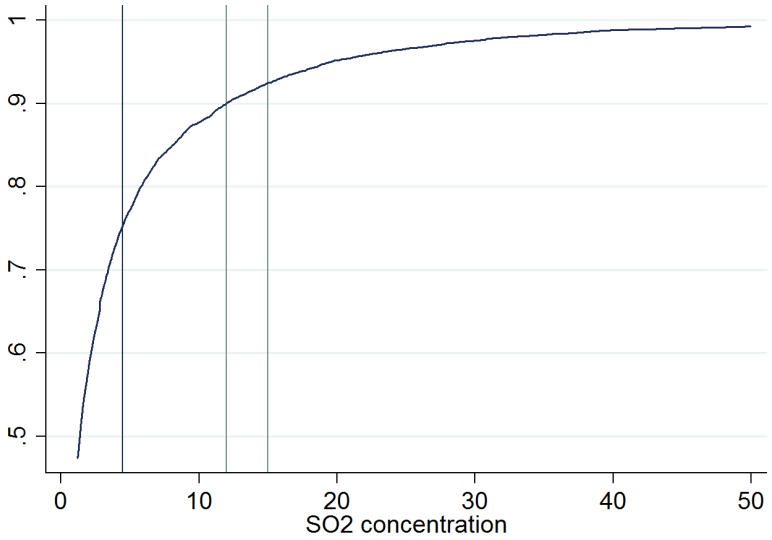
Notes: The figure plots the average decadal population growth rate for the period 1801-1891 in our sample cities.

Figure A3. Coal consumption in million tons, 1560–2001.



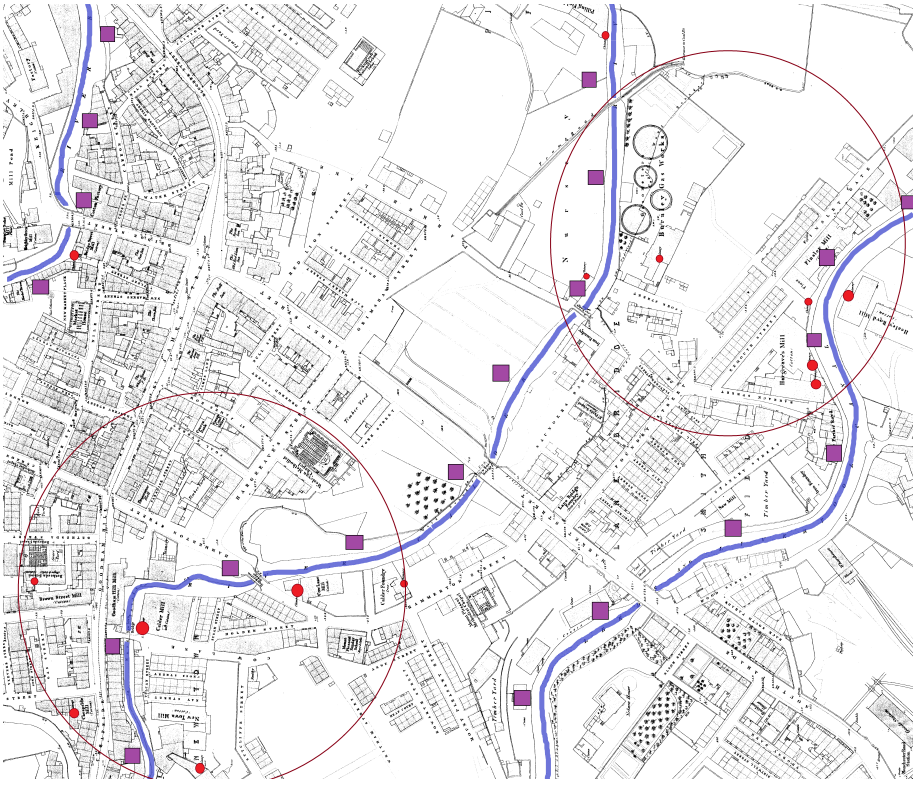
Notes: The figure illustrates the increase in coal consumption over the period 1560–2001. The figure is based on Warde (2007) who reports coal consumption in petajoule. To convert numbers from petajoule to tons, we use a conversion factor of 1:34,140. The two solid grey lines indicate the years 1817 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 A4. Cumulative of pollution in our sample of 10000 parishes and National Ambient Air Quality Standards (12-15 $\mu g/m^3$).



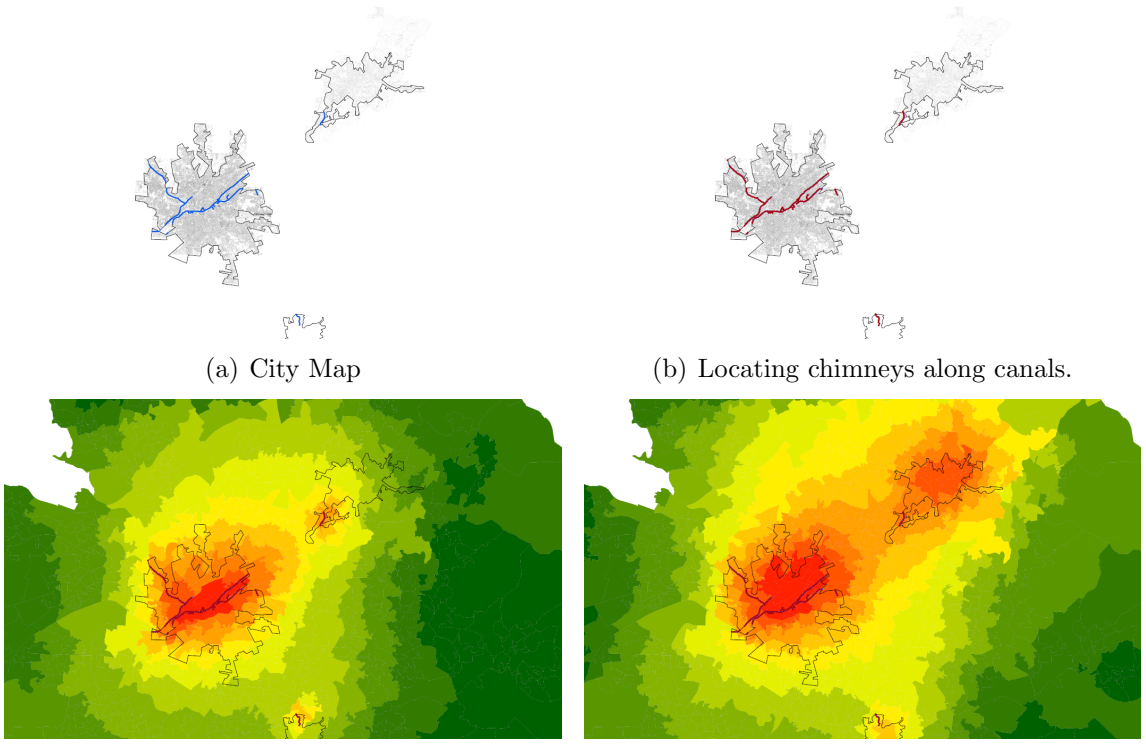
Sources: Authors' calculations.

Figure A5. Chimneys, canals, and synthetic chimney locations (example in Burnley).



Sources: Ordnance Survey Maps - 25 inch to the mile, 1842-1952. Chimneys are indicated with a red dot, and 1827 canals with blue lines. Synthetic chimneys are indicated with a purple square.

Figure A6. An illustration of the 2-stage empirical approach – correlation between pollution and deprivation.



(a) City Map

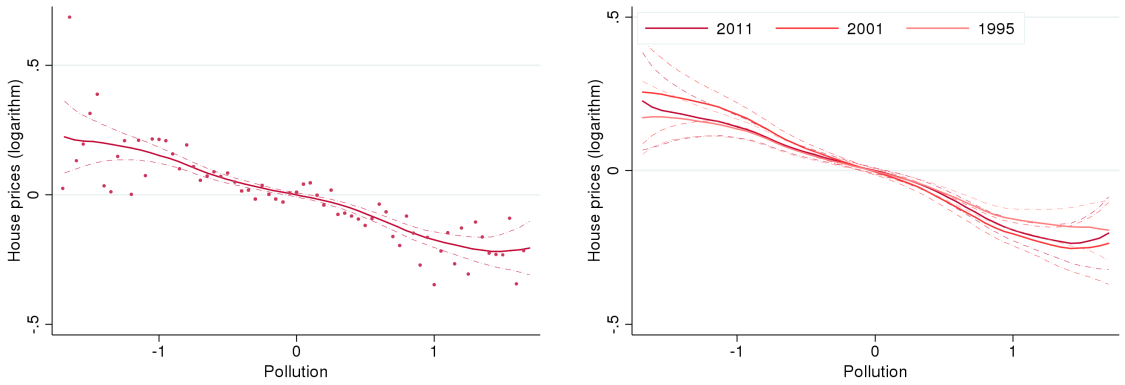
(b) Locating chimneys along canals.

(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 1817 canals with blue lines.

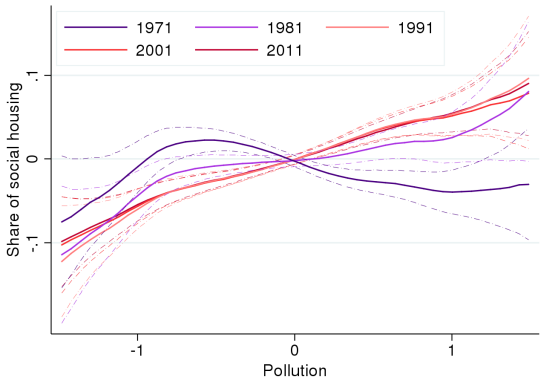
Figure A7. House transaction prices (y-axis) and pollution (x-axis) across neighborhoods – average and evolution between 1995 and 2011.



(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 A8. Social housing (y-axis) and pollution (x-axis) across neighborhoods in 1971, 1981, 1991, 2001 and 2011.



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

E Additional tables

Table A1. Pollution and shares of low-skilled workers in 1881 and 2011 – the role of covariates.

Share of low-skilled workers	1881	2011
Pollution	.0327	.0322
	(.0065)	(.0067)
Distance hall (inverse)	3.29	-2.92
	(3.26)	(3.53)
Share area (city)	.0084	-.0159
	(.0144)	(.0130)
Area	-.0030	-.0009
	(.0006)	(.0004)
Maximum elevation	.0001	-.0005
	(.0003)	(.0002)
Minimum elevation	-.0012	.0000
	(.0004)	(.0004)
Average elevation	.0002	-.0004
	(.0006)	(.0004)
Distance canals (inverse)	.0700	.0293
	(.0152)	(.0362)
Share low-skills (1817)	.1237	.0773
	(.072)	(.0747)
Share farmers (1817)	-.1479	-.0343
	(.0755)	(.0780)
Longitude	.0386	.0804
	(.0224)	(.0226)
Latitude	.0006	.0084
	(.0239)	(.0184)
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 A2. Pollution and shares of low-skilled workers – difference-in-difference specification (LSOA level).

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
Trends (city)	Yes	Yes	Yes	Yes
Trends (geography)	No	Yes	Yes	Yes
Trends (topography)	No	No	Yes	Yes
Trends (coordinates)	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. The basic specification is a panel specification with LSOA fixed-effects and city-specific trends. The set of geographic controls include distance to the city hall, share of LSOA within the city borders in 1890 and the area for the LSOA. The set of topographic controls include the average, maximum and minimum elevations for the LSOA and the distance to canals as of 1817.

Table A3. Pollution and shares of low-skilled workers – difference-in-difference specification (parish level).

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
Trends (city)	Yes	Yes	Yes	Yes
Trends (geography)	No	Yes	Yes	Yes
Trends (topography)	No	No	Yes	Yes
Trends (coordinates)	No	No	No	Yes

Robust standard errors are reported between parentheses. Each cell is the result of a separate regression. The unit of observation is a Lower Super Output Area/year. The basic specification is a panel specification with parish fixed-effects and city-specific trends. The set of geographic controls include distance to the city hall, share of parish within the city borders in 1890 and the area for the parish. The set of topographic controls include the average, maximum and minimum elevations for the parish and the distance to canals as of 1817.

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 canals as of 1817. 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. Pollution and deprivation measures in 2011.

Deprivation	Income	Empl.	Educ.	Health	Housing	Crime
Pollution	.0856	.0677	.0799	.0528	.0101	.0448
	(.0166)	(.0137)	(.0130)	(.0112)	(.0050)	(.0071)
	[.3012]	[.2420]	[.2741]	[.2019]	[.0337]	[.1817]
Observations	4,519	4,519	4,519	4,519	4,519	4,519
Fixed effects (city)	Yes	Yes	Yes	Yes	Yes	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. 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 canals as of 1817. The set of 1817 controls include the parish-level shares of farmers, managers and blue-collar workers. 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).

Table A6. Pollution, house prices and transactions (2000-2011).

	House prices		Transactions	
VARIABLES	(1)	(2)	(3)	(4)
Pollution	-.1067	-.0513	-.0781	-.1515
	(.0185)	(.0109)	(.0226)	(.0248)
	[-.1888]	[-.0908]	[-.1421]	[-.2757]
Observations	4,519	4,519	4,519	4,519
Fixed effects (city)	Yes	Yes	Yes	Yes
Controls (house types)	No	Yes	No	Yes
Controls (topography)	Yes	Yes	Yes	Yes
Controls (1817)	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 set of topographic controls include the average, maximum and minimum elevations for the LSOA and the distance to canals as of 1817. The dependent variables are the (logarithm of the) average house prices and number of transactions between 2000 and 2011. Controls for house types include the average shares of detached, semi-detached, terraced houses (flats/maisonnettes being the missing share) as well as new houses between 2000 and 2011.

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
Controls (geography)	Yes	Yes	Yes	Yes	Yes
Controls (topography)	Yes	Yes	Yes	Yes	Yes
Controls (coordinates)	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 topographic controls include the average, maximum and minimum elevations for the LSOA and the distance to canals as of 1817. 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 in 1881 and 2011 – controlling for 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
Fixed effects (city)	Yes	Yes	Yes	Yes
Controls (amenities 1881)	Yes	Yes	Yes	Yes
Controls (amenities 2011)	No	Yes	No	Yes
Controls (topography)	Yes	Yes	Yes	Yes
Controls (1817)	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 set of topographic controls include the average, maximum and minimum elevations for the LSOA and the distance to canals as of 1817. Controls for amenities in 1881 include the number of parks, schools, theaters, museums, churches, hospitals per 100 inhabitants at the LSOA level. Controls for amenities in 2011 include the number of parks, schools, theaters, museums, churches, hospitals, public buildings (e.g., town halls), courts, police stations, bus or train stations, botanical gardens, banks and conference centers per 100 inhabitants at the LSOA level.

Table A9. Pollution and shares of low-skilled workers in 2011 – controlling for building age.

Share of low-skilled workers	(1)	(2)	(3)
Pollution	.0354	.0322	.0294
	(.0074)	(.0081)	(.0088)
	[.2023]	[.1840]	[.1676]
Observations	4,519	3,478	1,982
Sample	All	After 1940 > 50%	After 1970 > 50%
Fixed effects (city)	Yes	Yes	Yes
Controls (building age 2011)	Yes	Yes	Yes
Controls (topography)	Yes	Yes	Yes
Controls (1817)	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 set of topographic controls include the average, maximum and minimum elevations for the LSOA and the distance to canals as of 1817. Building age controls include the shares of dwellings constructed before 1900, between 1900 and 1940, 1940 and 1970, 1970 and 2000, and after 2000. *After 1940* > 50% (resp. *After 1970* > 75%) only includes LSOA with more than 50% of dwellings constructed after 1940 (resp. 1970).

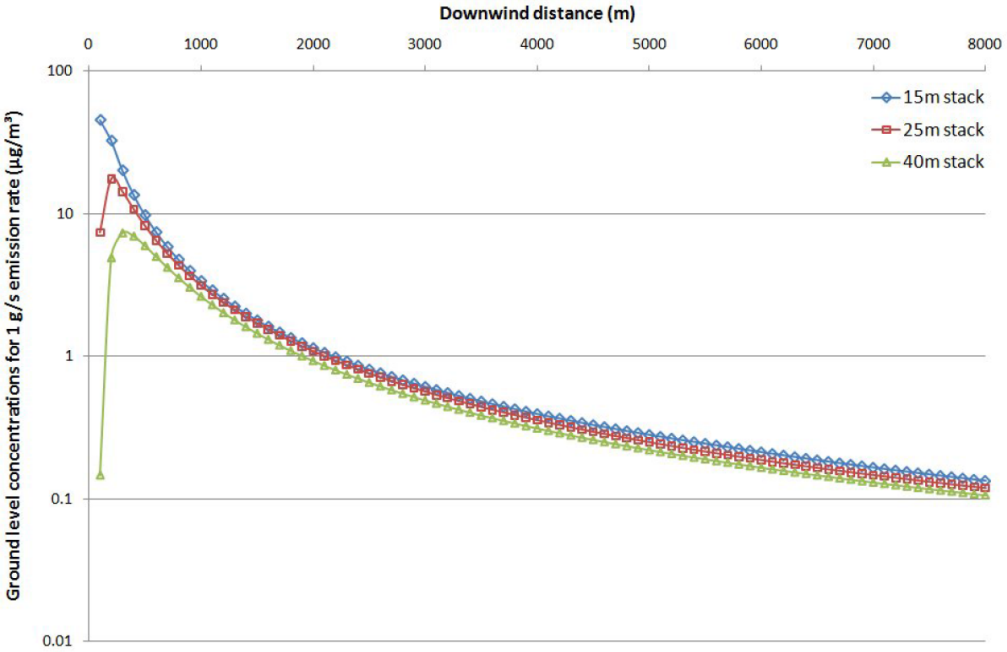
Table A10. Pollution and amenities in 1881 and 2011.

Effect of pollution on ...	1881	2011
Parks and recreation areas	.0000	.0444
	(.0001)	(.0167)
Schools and universities	-.0254	.0010
	(.0479)	(.0117)
Theaters, museums	-.0060	-.0053
	(.0040)	(.0095)
Churches	-.0361	.0196
	(.0536)	(.0095)
Hospitals	-.0074	-.0053
	(.0056)	(.0023)
National and local authorities		.0317
		(.0199)
Justice (courts and police stations)		.0111
		(.0095)
Banks/conference centers		-.0213
		(.0092)
Transport (bus and train stations)		.0346
		(.0119)
Botanical gardens and zoos		-.0085
		(.0041)
Observations	3,814	4,519
Fixed effects (city)	Yes	Yes
Controls (geography)	Yes	Yes
Controls (topography)	Yes	Yes
Controls (coordinates)	Yes	Yes

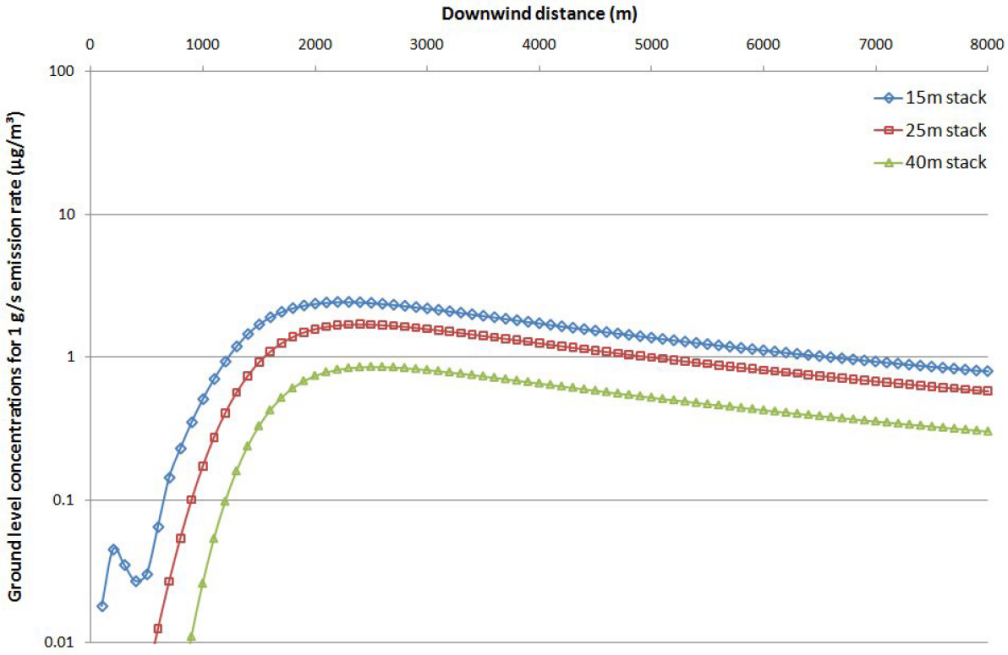
Standard errors are reported between parentheses and are clustered at the parish-level. Each coefficient is the estimate for pollution in a separate regression. The unit of observation is a Lower Super Output Area. The set of topographic controls include the average, maximum and minimum elevations for the LSOA and the distance to canals as of 1817. Amenities in 1881 and 2011 are computed per 100 inhabitants at the LSOA level.

F Sensitivity (ADMS 5)

Figure A9. Wind patterns and industrial air pollution – stable and unstable conditions.



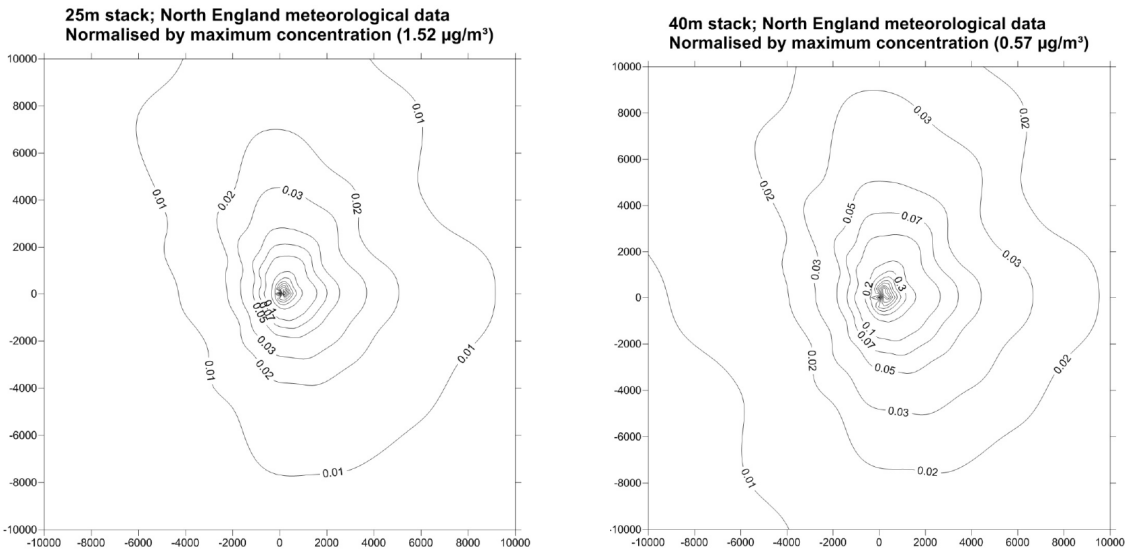
(a) Unstable conditions.



(b) Stable conditions.

Sources: ADMS 5.

Figure A10. Wind patterns and industrial air pollution (North England, flat terrain) – sensitivity to chimney heights.



(a) Chimney height: 25 m (benchmark).

(b) Chimney height: 40 m (benchmark).

Sources: ADMS 5.

G Descriptive statistics

Table A11. Air Pollution measures in the neighborhoods of Manchester.

	Deposits	Model estimates
Station	m. tons/ sq. m.	$\mu g/m^3$
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 A11. Cities .