Loss in the Time of Cholera: Long-run Impact of a Disease Epidemic on the Urban Landscape

Attila Ambrus, Erica Field and Robert Gonzalez*

October 27, 2015

Abstract

How do geographically concentrated income shocks influence the long-run spatial distribution of poverty within a city? We examine the impact on housing prices of a cholera epidemic in 19th century London in which one in seven families living in one neighborhood experienced the death of a wage earner. Ten years after the epidemic, housing prices are significantly lower just inside the catchment area of the water pump that transmitted the disease, despite being the same before the epidemic. Moreover, differences in housing prices persist and grow in magnitude over the following century. Census data reveal that price changes coincide with a sharp increase in population density at the border, consistent with anecdotes of impoverished residents taking in subtenants to make ends meet. To illustrate a mechanism through which idiosyncratic shocks to individuals that have no direct effect on infrastructure can have a permanent effect on housing prices, we build a model of a rental market with frictions, with poor tenants exerting a negative externality on their neighbors, in which a locally concentrated negative income shock can permanently change the tenant composition of the affected areas.

Keywords: Poverty dynamics, neighborhood externalities, real estate prices, health shocks

JEL Classification: O18, R21, R31, D62

---

[*] The authors are from Duke University (Field and Ambrus) and UNC-Chapel Hill (Gonzalez). We thank Patrick Bayer, Sarah Gelfand, and seminar participants at the University of Stockholm, Stockholm School of Economics, Northwestern, Harvard/MIT, SITE, and University of Michigan for useful feedback, and Volodymyr Baranovskyi and Veronica Polin for excellent research assistance. All errors are our own.

"Indeed, it is the peculiar nature of epidemic disease to create terrible urban carnage and leave almost no trace on the infrastructure of the city." [Steven Johnson, The Ghost Map, p.277]

1 Introduction

Can disease outbreaks exert a permanent effect on the geography of urban poverty? While it is well understood that illness is impoverishing, because health shocks have no direct impact on infrastructure or land, it is not obvious that epidemics that affect a small number of residents would leave an economic footprint on a city. As the quote above illustrates, a common presumption is that residential migration will preserve the spatial distribution of income in the long run, erasing such shocks from the map over time. In this manner, idiosyncratic income shocks to households should not lead to lasting "pockets of poverty" within a city. Yet, in reality, spatial discontinuities in the value of urban land are frequently observed and do not always appear to be related to discrete changes in local amenities.

We examine this question in the context of a cholera epidemic that hit a single urban parish in the metropolis of London in 1854. Over the course of one month, 660 residents living in the 0.5-mile radius of St. James Parish died from cholera, implying that roughly 5% of families were suddenly impoverished because of the loss of a potential wage earner. The outbreak was eventually attributed to contaminated groundwater that leached into one of the thirteen wells serving the parish. As a result, the impact was concentrated in one particular neighborhood of the parish wherein a much higher fraction of families – 13% – experienced a loss.

In this paper, we test empirically whether the likelihood of a residence being struck by cholera as determined by its location relative to the source of the 1854 epidemic is correlated with real estate prices soon afterwards and long after the epidemic ended. Although the research question can be generalized to any setting, there are two main reasons for focusing on this specific event. First, the 1854 outbreak in London provides a unique natural experiment that helps isolate the causal influence of a locally concentrated income shock on long-run outcomes. As illustrated in elegant detail by the 19th-century epidemiologist John Snow, since the outbreak was traced to one particular water pump, there are sharp changes in rates of cholera at the boundary of the catchment area for the contaminated well, where other property attributes such as access to public goods can be assumed to change smoothly. This makes it possible to isolate the causal effect of cholera exposure using a regression discontinuity framework.

A second important advantage is that very detailed microdata on deaths and property characteristics are available from this particular neighborhood at the time of the epidemic, which allows for careful examination of the identifying assumptions and causal mechanisms giving rise to changes in property prices post-epidemic.[1]

Our results reveal that houses inside the cholera-affected catchment area suffer a roughly 15% loss in rental value within a decade of the epidemic. More surprisingly, differences in property values persist for 150 years, and show no signs of convergence: In 1936 we estimate 37% lower property values just outside relative to just inside the catchment area. Differences remain significantly higher in contemporary real estate sales prices. Associated with the fall in rental prices is a significant increase in residential crowding inside but not outside the catchment area that emerges shortly after the epidemic. This pattern is consistent with historic accounts of rentors taking in short-term tenants when money is tight, which had the potential to produce significant block-level externalities in an already crowded neighborhood.

We make sense of these patterns by building a simple model of a landlord's rent-setting behavior in a rental market with rich and poor tenants and block-level externalities from living among the poor. In this setting, we consider what happens to the landlord's optimal strategy to attract or retain rich tenants when multiple tenants on his block simultaneously experience a negative income shock that transforms them from rich to poor, producing immediate block-level externalities. As we show in the paper, if the fraction of households on the block that transition to poverty is sufficiently high, it can become preferable for the landlord to set a rental price that attracts and retains poor tenants rather than offering a discount to rich tenants to live on a poor block that will only slowly transition back to rich.

Consistent with this story, we show that, inside the catchment area of the epidemic, houses that did not experience a death from cholera experience a change in rental value nearly as large as those that did. This suggests that the negative impact of cholera on household income reduces the value

---

[1] Not only did scientists and city officials investigating the outbreak collect detailed microdata from the parish at the time of the shock, but land tax assessment records from St. James are available for nearly every decade until the mid-20th century. In many parts of the city land tax records were not well-preserved.

of all properties within the neighborhood not only those that were hit by the disease. Moreover, the magnitude of the loss in rental value of a particular property depends fundamentally on the fraction of households in the immediate vicinity that experienced a cholera death. Data on migration between 1853 and 1864 indicate that the degree of neighborhood impoverishment encourages the unaffected to leave and the affected to stay, consistent with a story in which individuals get disutility from poor neighbors and an endogenous rental price response allows the newly poor to stay on a heavily hit block.

Our results tie into a vast literature in economic geography on long-run persistence of income differences across space. Particularly related to ours are a handful of papers that show evidence of persistent income differences across cities or towns even long after specific sources of economic advantage have become obsolete. For instance, Bleakley and Lin (2012) show how geographic features (portage) that contributed to economic activity historically but stopped mattering in the 19th century are correlated with the long-run income growth of cities. Similarly, Hanlon (2014) shows that a short-lived price shock to the cotton industry in 19th-century England is associated with the long-run economic growth of cities and towns that relied on cotton at the time of the shock despite the fact that the price of cotton rebounded within a decade. Of the same flavor is a paper by Dell (2010), which shows that the economic performance of towns in rural Peru is correlated with their position with respect to the boundary of a colonial labor tax catchment area that was abolished in the early 19th century.

Our paper builds on the above literature by undertaking a similar exercise within the microenvironment of one London parish. The contribution of this approach is that the central mechanisms for persistence that have been emphasized in previous work are not applicable within the parish setting, which constitutes a single economic and institutional environment. In particular, the first two papers cited above interpret their findings largely through changes in population growth that accompanied economic development or economic shocks, and the path dependence that demographic trends created via economies of scale in industry. The last paper interprets persistence through the lens of institutions, arguing that long-run differences in economic development inside and outside the catchment area for the colonial labor tax would not have occurred in the absence of persistent differences in local institutions.

Our setting is sufficiently small to preclude either interpretation - differences in property values within a single administrative district occupying merely 164 acres of land cannot possibly be attributed to differences in the evolution of local institutions or to a restructuring of economic activity in response to the disease epidemic. As a result, what our results add to the existing literature on persistence is evidence for an alternative means through which temporary and localized economic shocks can lead to long-term changes in the spatial distribution of poverty. As such, the mechanism we highlight is relevant for understanding urban ghetto formation more generally. Moreover, this particular channel of persistence may contribute to income divergence across local economies that has been documented in the literature.

In this sense, our findings suggest a broader set of channels related to residential sorting through which we might observe persistent effects of historic shocks on the long-run economic growth of neighborhoods or communities than is often considered in the literature. They also illustrate the potential economic cost of spatially correlated shocks when there are significant externalities from neighbors' socio-economic status, and in that manner tie into the large literature documenting and modeling neighborhood externalities in real estate values. Because the results suggest a potential source of misallocation of households across space, they also provide a rationale for third party interventions in real estate markets such as urban renewal projects frequently undertaken by municipal governments, as do results from existing work such as Hornbeck and Keniston (2014).

Our theoretical model is closely related to spatial models of location choice and segregation (Schelling (1969, 1971, 1978), Pancs and Vriend (2007)), but there are several key differences. First, agents in the above models follow simple behavioral rules, while in our model they are fully forward-looking utility maximizers. Second, in our model rent-setting landlords coordinate the movement of tenants in and out of the block. Third, instead of a self-contained city, our model features a block situated in an open world, where tenants can move in and out. The focus of our paper is also different: it is relatively easy to establish in our model that ultimately the block contains only one type of tenant, instead most of our focus is on characterizing the initial conditions under which the block converges to poverty.[2] On a technical level, our paper is related to asynchronous-move

---

[2] Mobius (2000) and Guerrieri et al. (2013) feature dynamic models of location choices, with market clearing rental prices, mainly focusing on the issue of how an inflow of new agents changes segregation in a city. These models do not feature price-setting landlords coordinating location choices of different types of agents, causing multiplicity of equilibria and limited predictive power regarding the long-term composition of a particular block of the city, which is the main focus of our analysis. A common assumption we share with Guerrieri et al. (2013) is that poor agents exert negative externality on their neighbors. Hornbeck and Keniston (2014) also investigate housing choices with externalities, but in a very different context, in which the qualities of neighboring buildings

dynamic games, and the role of asynchronicity of moves in coordination problems.[3]

The remainder of this paper proceeds as follows. Section 2 describes the study context and dataset. Section 3 describes the empirical strategy. Section 4 examines the immediate and long-term impacts of the cholera outbreak on housing prices. Section 5 provides a theoretical analysis of the channels of persistence, while other potential contributing factors are considered in Section 6. Lastly, Section 7 concludes.

2 Background

We study the evolution of property values of all residences in the London parish of St. James, Westminster, in the district of Soho, from immediately prior to the cholera epidemic of 1854 to more than a century after. Below we describe the setting and natural experiment, and the data sources utilized in the empirical investigation.

2.1 The Broad Street Cholera Outbreak of 1854

In 1854, St. James was a working-class neighborhood of 35,000 residents and a heavy commercial district that housed a large number of self-employed.[4] The most common occupation in the neighborhood was tailor, followed by shoemaker, domestic servant and mason. It was also the most crowded parish of London at the time, housing 432 people per acre. Density was high largely because it had been a previously wealthy neighborhood containing many multi-story buildings that became working class as the city expanded. On average, a single address in the neighborhood contained four families.

While it was crowded and economically diverse, St. James was not a particularly poor London neighborhood. As described by historian Steven Johnson, “By the [1850s], the neighborhood had turned itself into the kind of classic mixed-use economically diverse neighborhood that today’s “new urbanists” celebrate as the bedrock of successful cities: two-to-four story residential buildings with

---

affects the incentives of a house owner to invest in the quality of her house. This leads to very different long-term dynamics than in our paper, and in particular they show that a negative shock to collateral value can increase the quality of a neighborhood, by coordinating the owners’ investments during the rebuilding phase.

[3] Seminal papers in this literature include Farrell and Saloner (1985), Maskin and Tirole (1988a,b) and Lagunoff and Matsui (1997).

[4] Parish population estimate (35,406) from the 1851 UK Census. The population was nearly stationary between 1841 and 1861 (The Cholera Inquiry Committee, 1855).

storefronts at nearly every address, interlaced with the occasional larger commercial space. (...) The neighborhood's residents were a mix of the working poor and entrepreneurial middle-class." (Johnson, 2007, "The Ghost Map" p. 18).

The majority of occupants of St. James were renters (93%) with absentee landlords that owned multiple flats on the block. Rental contracts at this time mainly took the form of yearly tenancy agreements, meaning that either landlord or tenant could terminate or renegotiate the agreement without cause, though only when the 12-month lease expired (Beeton, 1861).[5] Legally-speaking, long-term leases were also possible, as were "tenancy-at-will" agreements, in which either party could leave at any time. Presumably, the latter were too costly to be desirable for absentee landlords and tenants alike. With respect to long leases, while we cannot observe the distribution of contract lengths in our sample, anecdotally 3-year and 5-year contracts were also common.[6] Thus, while tenants had weak rights by 20th-century UK standards, their tenancy rights were comparable to contemporary US standards.

In August 1854, St. James experienced a sudden outbreak of cholera when one of the thirteen shallow wells that serviced the parish – the Broad Street pump – became contaminated with cholera bacteria and remained so for weeks until groundwater gradually flushed it away.[7] At that time, the mode of cholera transmission was still unknown, so residents were unaware they should stop using the local water source in order to avoid infection. As a result, the bacteria quickly infected a large fraction of the parish population that lived within the catchment area of the Broad Street pump (henceforth BSP), which encompassed 57 densely packed blocks.[8]

The epidemic was fast and furious, and - because the source of transmission was stagnant rather than circulating - highly geographically concentrated. By the epidemic's close within a month, 660 residents of St. James had died from cholera, or 3% of the population. Within the BSP catchment

---

[5] Furthermore, landlords were required to notify tenants of changes to contract terms or eviction within 6 months of the end of the contract else tenants were entitled to one more year of occupancy under current terms.

[6] According to tenancy law at the time, "Where an annual rent is attached to the tenancy, in construction of law, a lease or agreement ... is a lease from year to year, and both landlord and tenant are entitled to notice before the tenancy can be determined by the other. (Beeton, 1861)" The fact that land tax data provide an annual rental amount for almost all properties in the sample suggests that most were not short-term lease agreements.

[7] Most of the historic details provided in this section come from Johnson (2007), a detailed account of the 1854 epidemic and its aftermath.

[8] The epidemic was later attributed to a leaking cesspit adjacent to the well. It was standard practice at the time to locate wells away from active cesspits, and the cesspit that caused the outbreak had in fact been out of use for several years since the parish had gained access to sewer lines. However, when a baby in St. James came down with cholera, her mother eventually made use of the inactive cesspit, causing bacteria from the initial victim to become trapped in the well below the Broad Street pump.

area, an estimated 16% of residents had contracted the disease and approximately 8% died.[9]

One particular public health authority, Dr. John Snow, had postulated after studying patterns of past cholera outbreaks that water may be the main transmission channel of cholera. In order to collect evidence to support his theory, he immediately began mapping the victims of the St. James outbreak alongside information on the location of wells within the district. He quickly saw a stark pattern of disease incidence in which nearly all victims were clustered around Broad Street pump. According to his diagrams, two-thirds of the residents of this tightly packed parish were at almost zero risk of contracting cholera because they happened to live closer to one of the twelve different water pumps in the parish.

Snow brought the data to health authorities and convinced them to disable the pump. Soon after the cholera epidemic subsided, government officials removed all old cesspits from the neighborhood and replaced the Broad Street pump handle. The epidemiological analysis conducted by John Snow provided the key evidence to prove the oral-fecal method of disease transmission, which fueled a long era of public health investment in water and sewerage infrastructure.

2.2 The Impact of Cholera on Neighborhood Poverty

In this paper we make use of the natural experiment provided by the swift and unanticipated cholera outbreak of 1854 to examine how geographically concentrated income shocks can influence the long-run spatial distribution of poverty within a neighborhood.

It is first worth considering the scale of the cholera epidemic within the Broad Street pump neighborhood, which contained approximately 500 properties and 2000 families.[10] By our estimates, 42% of all properties in the neighborhood experienced at least one cholera death during the course of the epidemic, some of which lost multiple residents: 19% of households lost just one member and 23% of households experienced the loss of more than one life, which implies that approximately 5%

---

[9] The reported mortality rate was around 50% at that time (The Cholera Inquiry Committee, 1855). The main reason that so many residents were spared from contagion is that a large number fled the neighborhood during the first few days to wait out the epidemic. Another limiting factor was the infrequency with which many residents drew water from the pump. Anecdotally it was common for households to take water only once every few days. Finally, heavy consumption of tea and alcohol, which have antimicrobial properties, protected many residents from exposure.

[10] Source: 1851 UK Census. According to tax records from 1853 there are 491 addresses that lie within the BSP catchment area. According to the 1851 census, the district contains 1650 addresses, 4439 families and 20,807 individuals.

of families lost one member and another 9% of families lost at least two members.[11]

Conceptually, we anticipate changes in the rental value of property arising out of the sudden impoverishment of residents inside the Broad Street pump catchment area. In particular, on the eve of the epidemic a large number of these families (and an even larger number of properties) were suddenly significantly worse off economically due to the death of a potential wage-earner. The available data do not detail which households experienced deaths of working-age members versus economic dependents, but aggregate figures published after the outbreak show that working-age adults were the most vulnerable to cholera transmission - 76% of deaths occurred among individuals aged 10 to 60 (Cholera Inquiry Committee, 1855). This implies that, in expectation, 13% of families inside the Broad Street pump catchment area and only 1% in the rest of St. James lost a wage-earner.[12] As a result of this tragedy, overnight the Broad Street area had become a neighborhood with many more destitute families than the rest of the parish: approximately one in seven families and two in five properties are likely to have transitioned suddenly from poor to destitute.

Not only were many households in the neighborhood hit by the epidemic, but because the area was densely packed, most residents lived on blocks that were heavily hit. In total, 80% of households in the Broad Street pump catchment area lived on a block in which at least 25% of residences had experienced a cholera death, and 25% lived on a block in which at least half of residences had experienced a death. The corresponding figures in the rest of the parish were 10% and less than 1%. The density of deaths within the neighborhood was relatively uniform, consistent with the assumption that nearly all residents relied on that pump for drinking water. Within 30 meters of the catchment area boundary, 73% of residents lived on a block in which at least a quarter of residences experienced a loss, relative to only 31% outside of the boundary.

We further presume that the death of a wage-earner leads very quickly to changes in household behavior that produce immediate, salient negative externalities on neighbors who live on or near

---

[11] These calculations are approximate because we lack data on the family membership of individual victims, though we know from aggregate population figures that the average residence in Soho housed three families and contained 13 members. In approximating the incidence of deaths across families, we assume that deaths are clustered within family,so we divide the total number of deaths recorded for a particular residence by two when the total number is under eight, and divide the total number of deaths by three when the number is under 12. Only one property experienced a number of deaths greater than 12 (18 deaths), which we estimate affected four families in the household.

[12] Author's calculations based on information on the age distribution of deaths published in the The Cholera Inquiry Committee (1855), the distribution of deaths across houses as collected by Cooper (1854), and the average number of families per house as recorded in the census of 1851.

the block. The forms of behavior change that are likely to create the largest and most immediate externalities in this setting are crowding of both people and animals. According to historical sources, it was common for tenants trying to make ends meet to take in additional tenants on short-term sub-lease. As noted by historian Sherwell (1897), "The announcement that may sometimes be seen in Soho of "Part of a room to let," represents what is frequently a very serious aggravation of the evils of overcrowding. In one case a small back-room was occupied by a young, newly-married couple, who took in a single-man lodger who slept in a chair. In the same house two back-rooms, both small, were occupied by a man and his wife and three men-lodgers, and the rooms were further let out at night for gambling purposes at the rate of one shilling per hour. Subsequently the woman (whose husband was a baker and therefore away all night) got rid of the men-lodgers and boarded a prostitute, and let her rooms out to this woman as a common brothel."

Anecdotally, it was also common for tenants even in the densest section of London to raise cash by crowding the apartment with farm animals and selling milk, eggs or dung. As detailed in Johnson (2007),"Residents converted traditional dwellings into "cow houses" - herding 25 or 30 cows into a single room (...) One man who lived on the upper floor of 38 Silver St. kept 27 dogs in a single room. He would leave (...) prodigious amounts of canine excrement to bake on the roof of the house"(Johnson, 2007, p.28).

It is safe to say that, particularly in such a tightly packed space, both more humans and more animals in neighboring homes would lead almost immediately to salient within-block externalities in the form of greater smell and noise, visible excrement (crowded sewers and cesspits, fewer street sweepers), disease and general misery (e.g. domestic violence, drunken brawls).[13]

2.3 Data Collection

To test whether property values respond to the outbreak, we gather several waves of data on housing prices of the roughly 1,700 housing units in St. James parish from 1853 to the present. This section describes the data used to define the catchment areas, real estate outcomes, and the baseline covariates employed in the analysis.

To determine the location of water pumps in Soho at the time of the cholera outbreak, we use

---

[13] Other sources of externalities are also possible, for instance higher exposure to crime, disease, lower public good contributions and sanitation, etc. For an empirical measurement of externalities among neighboring residents, see Rossi-Hansberg et al. (2010).

John Snow's cholera map (Snow, 1855), depicted in Figure 1a.[14] To track changes in real estate values, we construct a property-level panel database encompassing all residences in St. James that contains measures of property values for the years 1853 (pre-outbreak), 1864, 1936, 1995-2013, and 2015 obtained from three separate datasets. First, for the years 1853, 1864, and 1936, we collect data on the rental value and assessed land taxes from the National Land Tax Assessment records.[15] The Land Tax was first introduced in England in 1692 and formed the main source of government revenue until the late 19th century (Dowell, 1965).[16] Given their importance for public finance, coverage in tax records is very complete. In our study area, very few properties are missing from the records when compared to maps of the area. We also match the names of the primary occupant at each address across the 1853 and 1864 records to obtain a measure of residential turnover before and after the epidemic.

The Land Tax Assessment (and property taxes more generally) ended in 1963. Hence, for the years 1995-2013 we obtain property sales prices from the Land Registry of England (Land Registry, 2014). Records include the property address as well as the sale price and date of sale. We also obtain data on rental prices posted on the formal market for Soho over the past 5 years. Lastly, for the year 2015, we obtain house value estimates from *zoopla.co.uk*, UK's largest property listing website (Zoopla, 2015).

We digitized all valuations and addresses obtained from the records described above and employed two methods to geocode addresses. For historic records (1853, 1864, and 1936), we match addresses to detailed housing maps from the relevant time period.[17] For current house records, we geocode addresses using Google's geocoder tool.[18]

Next, to assess the spatial distribution of cholera deaths in and around the BSP area, we map

---

[14] Pump locations are also given on the Cholera Inquiry Commission map.

[15] We obtain these records through *Ancestry.com* (Ancestry.com, 2011). The dataset also contains information on whether a specific property had been exonerated. In 1798, the Land Tax Redemption Office was created under a registrar, and the Land Tax became a perpetual charge, which could be redeemed by the payment of a lump sum and landowners were thereby exonerated. The lump sum equaled 15 years tax, but the tax could be redeemed by purchasing 3 per cent consols in government stock which would yield an annuity exceeding the tax by a fifth.

[16] The Act specified that real estate (both buildings and land) were to be taxed permanently (beginning in 1798). It nominated for each borough and county in England and Wales local commissioners to supervise the assessments and local collection. Roughly every decade properties were assessed for tax value. Individual tax assessments were made based on the actual rental values of land, which were recorded by assessors.

[17] In the case of 1853 and 1864, we use the Metropolitan Commission of Sewers' 1854 housing map as base map (Cooper, 1854). We match 1936 addresses using the *England and Wales Ordnance Survey* map as base map (Ordnance Survey, 1951).

[18] To assess the quality of geocoding, we randomly selected 10 percent of the sample and manually checked the geocoded addresses using Google maps. All records matched perfectly.

the total number of deaths by house using the Cholera Inquiry Committee's 1855 map (Cholera Inquiry Committee, 1855). These data were gathered immediately after the epidemic as part of an epidemiological study into the mode of transmission of cholera. Snow and local chaplain Richard Whitehead conducted a census of the neighborhood in which all residences were visited and asked to report any deaths from cholera or diarrheal disease that had occurred over the past month. The map provides information on both house locations and number of deaths per house. Figure **??** presents a portion of this map. These deaths were later verified by the Commission using death certificates registered in the vital statistics database. Missing from the CIC map, which records 632 deaths, are 28 cholera deaths registered on death certificates from individuals who were not reported in the census.[19]

We also gather data on neighborhood amenities from the same maps. In particular, in investigating the outbreak to determine its origin, scientists and city planners constructed careful records of the location of cholera deaths in the neighborhood along with a wide set of neighborhood amenities and infrastructure in and around St. James parish. A particularly rich source of data was the map of the area constructed by Metropolitan Sewage Commissioner Edmund Cooper immediately after the epidemic as part of the "Report to the Metropolitan Commission of Sewers on the house-drainage in St. James, Westminster during the recent cholera outbreak". Cooper worked with an existing map of the neighborhoods sewer lines and residences, and added visual codes to indicate the location of cholera deaths and the site of the presumed and actual 17th century plague pit (believed to be a potential source of cholera at that time). As described by historians, the map was "superbly detailed: old and new sewer lines were documented with distinct markings; each gulley hole was represented by an icon on the map, along with ventilators and side entrances and the street number of every house in the parish" (Johnson, 2007, p.129).

In addition to the location of sewers and sewer vents, the map also contains the location of all 13 water pumps, public urinals, as well as neighborhood amenities including public squares, churches, police station, fire stations, theaters, banks, and primary school, to which we create measures of walking distance from each residence. From the same map we also calculate the distance to other neighborhood features that may influence housing prices, including the presumed location of a large plague pit in Soho (this location was later proved incorrect), and the distance to the center of Soho.

---

[19] Unfortunately we do not know the exact address of these 28 individuals, so must exclude them from the analysis.

We use a similarly detailed map from 1951 to assess the location of amenities pertaining to the 1936 properties. In both cases we digitized and geocoded map data and calculated distance measures to each residence in the dataset.

We collect data on population density and composition from individual records in the United Kingdom Population Census of 1851 and 1861. For each address in St. James, we record number of residents, number of families, and the country of origin of each family head. As with the land tax data, addresses are transcribed and geocoded.

Finally, we compare the socioeconomic status of households across the BSP boundary about 45 years after the outbreak using Charles Booth's 1899 poverty survey of London (Booth, 1902). The survey contains household-level socioeconomic status obtained from a combination of interviews with London School Board visitors and existing records collected by Board visitors during their yearly home inspections.[20] We transcribe these data and then geocode houses based on the house number and street name provided in the survey and create four indicators of household-level socioeconomic status (very poor, poor, working poor, and middle class) from the recorded social class classifications.[21]

3 Empirical Strategy

Property-level data allow us to assess the change in rental value of properties located inside versus outside of the Broad Street pump catchment area where the cholera epidemic was concentrated. In particular, we employ a regression discontinuity (RD) design that makes use of the change in disease rates that occurred at the boundary of the catchment area.

---

[20] Passage of the Education Act in 1870 and further legislation in 1876 to enforce primary school attendance led to the creation of School Board visitors who were in charge of visiting households, investigating instances of non-attendance, etc. Part of their obligations was to keep records of all households in their districts and to update them annually (Weiner (1994), p 31). These sets of records constitute the primary source of Charles Booth's survey of London.

[21] We follow O'day and Englander (1993) when translating Booth's poverty classifications into the four socioeconomic categories used in our analysis. Specifically, we classify categories A and B as "Very Poor", C and D as "Poor", E as "Working Poor", and F and G as "Middle Class". Refer to Appendix Table B6 for the original eight-tier classification used by Charles Booth.

3.1 Catchment Area Boundary

As originally proposed by Snow, we define the catchment area according to a network Voronoi diagram of Soho, in which each of the 13 water pumps defines a point and the cells are determined according to the walking distance to each point. We calculate the shortest travel distance by plotting the wells on a georeferenced 1854 street map. Hence, the catchment area encompasses all residents for whom the Broad Street pump was the water source in closest walking distance.[22] A picture of the map we employ in our analysis is shown in Figure 1a, alongside the original boundary mapping of Snow in Figure 2a. Figure 1b depicts the catchment areas for all pumps where each dot indicates the location of the water pumps at the time of the outbreak, and the BSP catchment area is outlined in red.[23] In both figures, cholera deaths are marked by black bars, and a portion of the map is enlarged for clarity in Figure 3.

As is evident from both maps, the constructed catchment areas map closely with the spatial pattern of deaths from cholera. According to our calculations, 76% of cholera deaths occurred within the catchment area, which is close to the figures calculated by Snow. Contemporary accounts also suggest the existence of a discontinuity in cholera deaths at the Broad Street pump boundary. For instance, John Snow himself stated that "deaths either very much diminished, or ceased altogether, at every point where it becomes decidedly nearer to send to another pump than to the one in Broad Street"(Snow, 1855). Figure 5a plots the share of houses with at least one cholera death and Figure 5b plots the share of houses for which at least 25% of houses on the block experienced a casualty, providing further confirmation of this pattern.

This is a particularly striking pattern given that the boundaries do not determine actual assignment to a particular water pump, but merely delineate the likelihood of utilizing a particular pump. Some residents just outside the boundary may get water from the pump because it is convenient or preferred for some reason. For instance, Broad Street pump was located close to the local primary

---

[22] For a formal definition of network Voronoi diagrams refer to Erwig (2000), Okabe et al. (2000). For a definition applied to the John Snow's cholera map, refer to Shiode (2012). We determine catchment areas using the *Closest Facility* solver in ArcGIS Network Analyst.

[23] Following previous literature (e.g., Shiode (2012)) and John Snow's own accounts, we discard the pump located on Little Marlborough St. when constructing the catchment areas. In his cholera report, John Snow states that "the water of the pump in Marlborough Street, at the end of Carnaby Street, was so impure that many people avoided using it. And I found that the persons who died near this pump in the beginning of September, had water from the Broad Street Pump" (Snow, 1855)

school, and anecdotally children drank from the pump on their way to school.[24]

A few discrepancies between the two maps in Figure 2 merit explanation. First, there are two regions of the map in which our calculation of minimum walking distance differs from that of Snow. In the lower right-hand corner of St. James, Snow assigns a cluster of houses on Berwick St. to the Broad Street pump catchment area, when they are clearly closer to the Rupert St. pump (on both his map and the Cooper map). While it is true that a disproportionate number of deaths occurred in these houses, there is no recorded information on residents of this neighborhood favoring the Broad Street pump, so we choose to include them in the catchment area. Second, based on numerous historical accounts, the pump at Little Marlborough St. was very rarely used for drinking water because of the foul smell and taste of sulphur. As Snow and Whitehead documented in interviews with residents of houses located closest to this pump, because Broad Street pump was less than 100 meters from Little Marlborough pump, they preferred to get drinking water from Broad Street, a fact Snow uncovered in seeking to explain why so many residents of the area became infected with cholera when the Marlborough pump was not contaminated. Given this, we exclude the Little Marlborough pump from among the points on the Voronoi map. Since it is possible that Snow had better information on actual walking distance than we do now, in Section 4 we verify that our estimates are robust to those using the catchment area as drawn by Snow.

3.2 RD Specification

We use a spatial regression discontinuity (RD) design that takes advantage of the discontinuity in deaths caused by well access to estimate the effect of cholera on real estate outcomes. We present results using two approaches typically followed in the spatial RD literature. First, we exploit the two-dimensional nature of the BSP boundary to estimate *conditional treatment effects* at various points along the treatment boundary following Imbens and Zajonc (2011).[25] Second, we follow the usual approach in the literature by specifying a one-dimensional forcing variable, namely

---

[24] It is also possible that, after the onset of the epidemic, there may have been sources of secondary infection of an indeterminate location. That is, since the disease is spread through fecal-oral pathogens, waste from cholera victims could have infected others through the standard channels of diarrheal disease transmission (e.g. dirty hand of individuals tending to victims). However, cholera is hard to catch since the bacteria must be ingested so sources of infection are likely to have been limited to within the same household where water in storage may become contaminated.

[25] Although there are multiple studies exploring RD methods with a multidimensional forcing variable (e.g. Reardon and Robinson (2010), Wong et al. (2012), Keele and Titiunik (2013)), we mostly follow the notation and terminology in Imbens and Zajonc (2011).

the distance to the closest point in the BSP boundary.[26] This is the equivalent of subtracting the cutoff value from the forcing variable in the one-dimensional design and then using this transformed forcing variable to estimate a single, boundary-wide average effect.

Broadly speaking, the first approach estimates treatment effects using observations within a neighborhood of a specific point in the treatment boundary. This exercise is then repeated for various points along this boundary thus providing a distribution of these effects along this dimension. Formally, let $\mathbf{b}_j$ with $j = 1, \ldots, J$ denote the coordinate vector of point j on the BSP boundary and let $N_h(\mathbf{b}_j)$ denote a neighborhood of size h meters around this point. To obtain the impact of cholera exposure on property values at boundary point $\mathbf{b}_j$, we use local polynomial regression. Specifically, we estimate the equation below for properties within neighborhood $N_h(\mathbf{b}_j)$:

$$y_{it} = \alpha + \gamma BSP_i + f(X_i) + \mathbf{W}'_{it}\beta + \epsilon_{it} \qquad \text{for} \quad X_i < h \tag{1}$$

where y_{it} is a measure of property i's value in year t; BSP_i is an indicator equal to 1 if property i falls inside the BSP catchment area; X_i is the distance in meters between property i and point $\mathbf{b}_j$ on the BSP boundary; and $f(.)$ is a polynomial of order K with $f(X_i) = \sum_{k=1}^{K} \delta^k X_i^k$ where the optimal choice of K is determined using Akaike's criterion as in Black et al. (2007) and suggested in Lee and Lemieux (2010). While baseline covariates are not needed for identification in the RD setup, they can improve the precision of the estimates (e.g., Lee (2008), Imbens and Lemieux (2008)), therefore we include vector $\mathbf{W}_{it}$ of property and street level characteristics in year t. Table 1 provides summary statistics for these covariates in the pre-outbreak period. Bandwidth h is chosen optimally following Imbens and Kalyanaraman (2012).[27] In the above equation, γ is the causal effect of exposure to cholera on y_{it} at point $\mathbf{b}_j$.[28]

As discussed in Imbens and Zajonc (2011) and denoting the estimate of the conditional treat-

---

[26] See Holmes (1998), Black (1999), Kane et al. (2006), Lalive (2008), and Dell (2010) for examples of papers employing an RD design with distance to the treatment threshold as the forcing variable.

[27] In the local RD setting, our choice of bandwidth can be interpreted as using a rectangular kernel with bandwidth h. Although some studies suggest and use a triangular kernel (e.g., Fan and Gijbels (1996), Imbens and Zajonc (2011), Keele and Titiunik (2013)), our choice of a simple rectangular kernel is for practical purposes. Lee and Lemieux (2010) state that, in the RD setting, kernel choice has little impact in practice therefore simple kernels (i.e., rectangular) can be used for convenience.

[28] Note that the method used in this paper differs from Imbens and Zajonc (2011) in two aspects. First, we use local polynomial regression instead of local linear regression. Second, we do not specify any kernel function to weight observations near the treatment boundary. This is primarily done for convenience since, in the RD setting, kernel choice has little impact on results (Lee and Lemieux, 2010).

ment effect at point $\mathbf{b}_j$ as $\hat{\gamma}(\mathbf{b}_j)$, we estimate a boundary average effect γ as:

$$\hat{\gamma} = \frac{\sum_{j=1}^{J} \hat{\gamma}(\mathbf{b}_j) \cdot \hat{g}(\mathbf{b}_j)}{\sum_{j=1}^{J} \hat{g}(\mathbf{b}_j)} \tag{2}$$

where $\hat{g}(.)$ is the estimated bivariate density of properties' coordinate vectors evaluated at boundary points $\mathbf{b}_j$ and each $\hat{\gamma}(\mathbf{b}_j)$ is estimated using Equation (1) above within neighborhood $N_h(\mathbf{b}_j)$.[29] Letting $\mathbf{s} \in \mathbf{S}$ denote the coordinate vector of a property and $\mathbb{B}$ denote the set of all boundary points (i.e., not only the J points used in estimation), the expression above provides an estimate of the average effect γ given by $\int_{\mathbf{s}\in\mathbb{B}} \gamma(\mathbf{s}) g(\mathbf{s} \mid \mathbf{S} \in \mathbb{B}) d\mathbf{s} = \frac{\int_{\mathbf{s}\in\mathbb{B}} \gamma(\mathbf{s}) \cdot g(\mathbf{s}) d\mathbf{s}}{\int_{\mathbf{s}\in\mathbb{B}} g(\mathbf{s}) d\mathbf{s}}$. In subsequent discussions of results, we refer to the estimate in Equation 2 as the *averaged conditional treatment effects*.

Following Imbens and Zajonc (2011), we choose a number of evenly spaced boundary points $\mathbf{b}_j$ that cover the boundary reasonably well. We initially choose 80 boundary points but to assess the sensitivity of the results we repeat the exercise using half the number of points $\mathbf{b}_j$. To satisfy the Boundary Positivity assumption (Imbens and Zajonc, 2011), we restrict the analysis to neighborhoods $N_h(\mathbf{b}_j)$ with at least one property on each side and within close distance of the treatment boundary.[30] After these restrictions and given the spatial distribution of our 1853-64, 1936 and 1995-2013 samples along the BSP boundary, we are able to estimate conditional treatment effects for 47, 25, and 26 boundary points $\mathbf{b}_j$, respectively, out of the 80 points drawn initially. For illustration, Figure 11 presents the set of boundary points $\mathbf{b}_j$ for which we are able to estimate conditional treatment effects using the 1853-1864, 1936, and current house values samples.

For the case where we follow the one-dimensional approach, we estimate Equation (1) for all properties within distance h of the entire BSP boundary rather than for specific neighborhoods $N_h(\mathbf{b}_j)$ along this boundary.

---

[29] We estimate the bivariate density via kernel density estimation using the *Kernel Density* tool in ArcGIS Spatial Analyst.

[30] Boundary Positivity requires the existence of observations near the boundary in order to identify the treatment effect in the multidimensional RD setting. More specifically, letting $\mathbf{L}_i$ denote the latitude and longitude of property i, Boundary Positivity requires that for all $\mathbf{b}_j$ and $\epsilon > 0$, there are properties for which $P(\mathbf{L}_i \in N_h(\mathbf{b}_j)) > 0$. In the estimation, we only use neighborhoods for which there is at least one property within 15 to 20 meters (depending on the specification) on each side of the treatment boundary. The choice of 15 to 20 meters corresponds to about one-quarter of the optimal bandwidth h for the corresponding specification. Authors can provide results using different cutoffs upon request.

3.3 Validity of RD design

Identification of the treatment effect in Equation (1) requires that potential outcome functions $E[v_i(1)|X_i]$ and $E[v_i(0)|X_i]$, where $v_i(1)$ and $v_i(0)$ denote the outcome under treatment and control, respectively, must be continuous at the treatment boundary. Broadly speaking, the assumption implies that all property characteristics (i.e., determinants of v_i) must be a continuous function of distance to the BSP boundary. This allows properties that are geographically close to the BSP boundary to serve as plausible counterfactuals for similar properties inside the BSP area.

We test the validity of this assumption by examining the similarity across the boundary of neighborhood features in the year prior to the epidemic, including rental prices and property tax assessments recorded in 1853, and neighborhood amenities measured at the time of the epidemic. Table 1 compares various property characteristics inside and outside of BSP boundary. Columns (1) and (2) provide mean characteristics for the full estimation sample, Columns (4) and (5) provide the same information for properties within 100 meters of the boundary, and Columns (3) and (6) present the standard error for the difference in means for their respective specifications clustered at the block level.[31] Although there are significant differences in means across all properties, when we consider only properties within a 100-meter bandwidth – the largest we consider in the empirical analysis –, the difference in characteristics between catchment area and non-catchment area properties decreases in magnitude and many become statistically insignificant.

For a better depiction of the continuity of baseline covariates across the BSP boundary, refer to Figure 7 which presents averages for continuous 20-meter distance bins. To provide a more meaningful assessment of the continuity assumption, Columns (7) and (8) in Table 1 present the coefficients and robust standard errors from the estimation of a modification of Equation (1) using each variable in Table 1 as the dependent variable.[32] For comparability, all estimations in Column (7) use the same bandwidth (27.5 meters).[33]

Table 2 verifies that rental prices in 1853 are smooth at the boundary in parametric RD specifi-

---

[31] For reference, Table B1 provides Conley (1999) standard errors adjusted for spatial correlation along the block level clustered standard errors. Standard errors using either method do not vary significantly.

[32] More specifically, Column (7) presents the estimate of γ in $w_{it} = \alpha + \gamma BSP_i + f(X_i) + \epsilon_{it} \quad \text{for} \quad X_i < h$, where w_{it} is a baseline covariate in Table 1. For the RD polynomial, we use a local linear regression specification where $f(X_i) = \rho X_i + \phi BSP_i * X_i$

[33] Using a different outcome variable yields a different optimal bandwidth in each estimation, however, for comparability among all variables, we present the results using the minimum of all optimal bandwidths.

cations that condition on a third-order polynomial in distance to boundary rather than local linear regressions, with standard errors clustered at the block level.[34] The regressions in Table 2 differ with respect to the bandwidth chosen, control variables included, and the level of clustering. In Column (2), the sample is restricted to observations that fall within the optimal bandwidth (28) chosen as suggested by Imbens and Kalyanaraman (2012), and includes controls for all covariates listed in Table 1 and standard errors clustered at the block level.

We also alter the bandwidth to a narrower (Column (3)) and a wider margin (not shown) to test the sensitivity of our results to excluding and including observations farther away from the border. The bandwidths chosen for the narrow and wide specification reflect the narrowest and widest bandwidths in which our estimates remain statistically robust, although it is worth noting that the point estimates are similar in magnitude under a broader range of bandwidth values. In Column (4) we verify that our estimates are robust to a larger level of clustering (street) and in Column (5) we include fixed effects for the different segments along the boundary.

Overall, Table 1 and 2 reveal a robust absence of differences between houses inside and outside the BSP boundary with respect to real estate values. No differences in rental prices the year prior to the epidemic emerge when we control for all possible covariates, nor when we alter the bandwidth to be as narrow as 24 meters nor as wide as 70, which encompasses 92% of the Broad Street pump catchment area. Results are also robust to clustering standard errors by street, as well as controls for categories of street width (not shown). The patterns in Table 1 indicate that, prior to the outbreak, properties on either side of the BSP boundary are very similar: for 18 out of the 19 baseline characteristics the RD coefficients are statistically insignificant. More importantly, measures of property value (i.e., rental prices, assessed and exonerated taxes) do not differ across the boundary, so significant differences between properties on either side of the boundary in the post-outbreak period cannot be attributed to pre-existing differences.

Identification of the treatment effect in Equations (1) and (2) also requires that there be no endogenous sorting of properties (and individuals) within a close window around the BSP boundary. Support for this assumption comes from the fact that the BSP boundary is solely determined by whether properties have better access to the water pump on Broad Street relative to the other

---

[34] Optimal polynomial order of three was determined using Akaike's criterion as detailed in Black et al. (2007) and suggested in Lee and Lemieux (2010).

pumps in Soho at the time.[35] Thus considering that there is no obvious distinction between pumps and that water is a relatively homogeneous good, there is no clear incentive for properties or individuals to sort near the boundary of the catchment area of a specific pump.

In order to provide a more quantitative assessment on the validity of this assumption, Figure 6a shows a histogram of the forcing variable (distance to the BSP boundary) that uses 15-meter bins. "Negative" distances represent the distances of properties outside the BSP area. Note that there is no clear evidence of a jump in the density of properties across the treatment boundary (represented by the solid line at zero). Additionally, we perform McCrary (2008) test for breaks in the density of the forcing variable, as shown in Figure 6b. Similarly, the density does not change discontinuously across the boundary suggesting that, for a narrow window around the BSP boundary, there seems to be no endogenous sorting of properties.

The pattern we observe in the RD plots and regression estimates is consistent with the less rigorous assessment of those investigating the cholera epidemic at the time of the outbreak, who hailed the pattern of contamination as proving the role of the water pump by virtue of the fact that residents served by various pumps in Soho were otherwise of similar socio-economic background. In particular, unlike previous cholera outbreaks in the city that struck relatively poor districts, John Whitehead detailed in his report to the Cholera Inquiry Committee (1855) that the pattern of contagion in Soho was not related to the socio-economic status of residents: "What class of persons did the disease principally destroy? (...) it attacked and destroyed all sorts and classes alike" (Whitehead, 1854, p.7).

4 Results

We start by examining how exposure to cholera influences real estate prices and residential mobility in the short run (Tables 4 to 7), and then investigate how valuations evolve over the next century (Tables 9 and 10).

---

[35] "Better access" in this context refers to shorter walking distance.

4.1 First stage

Table B2 shows the estimated association between living within the catchment area and exposure to cholera. Since we have no reliable measures of cholera exposure other than death reports (e.g. incidence of individuals exhibiting cholera symptoms), we gauge the strength of our first stage by employing the RD framework of Table B2 where the outcome is deaths from cholera at the house level, the block level, and the neighborhood level. Our definition of neighborhood encompasses the block on which the residence is located along with all contiguous blocks, so contains anywhere from 2 to 7 blocks (mean number of blocks is 3.5). We look at both the fraction of households that experienced a cholera death and the number of deaths divided by the number of residences.

The results indicate a roughly doubling in the death rate inside and outside of the catchment area within a 31-meter radius of the boundary. While outside the boundary, 19% of households experience a loss from cholera, the rate is 32% inside the boundary. As described earlier, the contrast is even starker when we consider a dummy indicator of living on a block in which more than a quarter of households are affected by cholera: Inside the boundary 73% of households live on a heavily hit block, while outside the boundary only 30% do. Even within only 20 meters of the boundary, the rate of being heavily hit approximately doubles in and out of the catchment area.

4.2 Rental prices

Table 4 presents estimates of the effect of the epidemic on the evolution of property values a decade afterwards. The outcome variable in the first set of estimates is the change in (the logarithm of) the rental price of a particular property between 1853 and 1864, and the second set of estimates looks simply at (the logarithm of) the rental values in 1864 as an outcome. While the first outcome presumably increases precision by accounting for time-invariant property characteristics, there are several observations in the dataset that cannot be matched across years and are therefore excluded from those estimates. In particular, 41 properties (2%) with value data in 1864 have no corresponding data in 1853, and 52 properties (3%) with value data in 1853 have no data in 1864. While we suspect that the majority of these residences belong in the panel but cannot be matched due to address errors in the tax data, we cannot rule out property creation and destruction across years.

As shown in Columns (1) and (2) of Table 4, properties inside the boundary experience a 10-

13% loss in rental value in the decade after the epidemic. There is little difference between the one-dimensional and two-dimensional RD specifications. The pattern is almost identical when we look simply at rental price in 1864 as the outcome, which allows us to include all observations available in the 1864 dataset. We estimate a difference in rental prices at the boundary of 13% at the optimal bandwidth of 51, and the estimates change very little when we narrow the bandwidth to 24 (Column (5)) or widen it to 71 (not shown), cluster standard errors at the street rather than block level (Column (6)), or reduce the set of covariates to only the three most important predictors of property values in the neighborhood - whether the property has no sewer access or old sewer access, and distance to the closest public urinal (Column (7)).[36]

In Table B5, we show the same regression estimates using the assessed tax burden of the property in place of the rental value, which allows us to include an additional 325 properties (19%) that were either exonerated from taxation or owner-occupied, so have no recorded rental values. Including these additional observations has no effect on the estimates, in part because the rate of exoneration and owner-occupancy is particularly low near the boundary, leading to the inclusion of only 44 additional observations to the main regression estimates.[37] Tax assessments of exonerated properties presumably reflect the rate as it was assessed at the time of exoneration, although according to interpretations of Historians, the Land Tax Assessment Database remains vague on this matter. Because of the uncertainty in the construction of the valuations, our preferred specification uses only rental prices as outcomes.

4.3 Migration

What effect did cholera have on migration out of the neighborhood? Tables 5 and 6 look at linear probably estimates of residential turnover as a function of location inside the Broad St. pump catchment area. An important caveat is that our measure of migration captures only whether the primary occupant of the residence is recorded as having the same last name in 1853 and 1864, so does not capture the rate of migration of all families or individuals in the residence. However, deaths from cholera were described as being equally distributed among primary occupants and sub-leasing

---

[36] It is also the case that estimated magnitudes are similar when employing a cubic polynomial specification or a more flexible specification in which the distance function is allowed to take on a different shape on the two sides of the boundary (Appendix Table X).

[37] For unclear reasons, exonerated properties are heavily concentrated in the upper west corner of St. James, which lies outside of even the wide bandwidth we consider in the analysis.

tenants (as measured by occupancy on the first floor or upper floors of the home) (Whitehead, 1854).

Overall, the neighborhood is one of relatively high turnover: 56% of primary tenants of St. James parish changed residence between 1853 and 1864, and the rate is 65% within the Broad Street pump catchment area. However, at the boundary of the catchment area the difference is minimal. As shown in Column (1) of Table 5, the difference in mobility is only 7.6% within 44 meters of the boundary. When we increase the bandwidth to 87 meters, the difference is 11% and strongly significant as shown in Column (2).

In Columns (3) and (4), we consider the wider bandwidth in order to investigate the relationship between differences in mobility and experiences with cholera. As shown in Column (3), households that were hit by cholera were no more likely to leave their residence – households inside the catchment area that escaped cholera were also 10% more likely to move. However, as shown in Column (4), the difference in mobility inside and outside of the border is fully accounted for by the number of deaths experienced by nearby households (controlling for the total number of houses in the neighborhood). We see the same pattern in rental prices (Columns (5) and (6)): The decrease in rental prices is not explained by the houses that experienced a cholera death - surrounding houses fall equally in value. However, the rental price difference at the boundary *is* explained by the number of deaths that occurred at the neighborhood level.

In Table 6 we look at the interaction between household deaths and neighborhood deaths on the propensity to move out to better understand the impact of neighborhood impoverishment on the relocation incentives of those that became impoverished by the epidemic versus those that escaped it. Columns (1) and (2) consider all residences in the parish, while Columns (3) and (4) look only within the Broad Street pump area where the epidemic was concentrated. As shown in the second and third row of Columns (3) and (4), the greater the number of deaths in the neighborhood (excluding own death) the more likely are those that escaped cholera to leave, whereas the opposite is true for those who experienced a death (the sum of coefficient estimates in rows 2 and 3 is negative). Among cholera victims, the more houses in the neighborhood that were hit by the epidemic, the more likely they are to stay.

We interpret these patterns as evidence of neighborhood externalities having two distinct effects on neighborhood composition: On the one hand, externalities from poor neighbors (those who lost a household member to cholera) drive people away, presumably for the reasons discussed in the previous section. But why would they encourage the newly impoverished to stay? The most obvious explanation is that rental price responses to neighborhood impoverishment increased the affordability of housing in affected areas.

From the migration data we can gauge the degree to which neighborhood impoverishment from cholera was still a contributing factor in 1864. While turnover within the BSP neighborhood was higher than it was just outside the catchment area, it is also the case that many households that experienced a cholera death stayed in the neighborhood. In total, one third of households that experienced a cholera death were recorded as living in the same residence ten years later.[38] The continued presence of households that experienced a shock certainly had an effect on the overall poverty of the block in much of the catchment area: Among a third of households in the BSP area, at least 15% of block residents were households that had been hit by cholera and stayed in the neighborhood. Furthermore, it is likely that many more of those who experienced a negative socio-economic shock as a result of the epidemic stayed on for part of the decade.

4.4 Residential Crowding

We use census data from just before (1851, Panel A) and soon after (1861, Panel B) the epidemic to more directly examine sorting of residents by socio-economic status onto blocks inside and outside of the boundary after the epidemic. Two relevant pieces of information are available in the census: the number of inhabitants within each dwelling, and the country of origin of inhabitants. We are interested in the first as a measure of crowding, one of the primary channels through which block-level externalities are believed to operate. Immigrant status is also a reasonable proxy for socio-economic status in this setting.

Results from these estimates are presented in Table 7. In Column (1), we see that population density is smooth at the BSP boundary in 1851, but in 1861 a sharp increase in population density emerges within the affected catchment area. In 1861, dwellings inside the boundary have 34% more people than those outside the boundary, or roughly five additional residents. Furthermore, the increase in density is accompanied by an increase in the proportion of immigrant families residing

---

[38] While high, this rate of residential movement is not substantially higher than the regular turnover rate in the neighborhood: Among households that were neither hit by cholera nor living within the cholera-affected area (BSP), only 48% resided at the same address after a decade.

in the dwelling. In 1861, there are nearly twice as many immigrant families inside relative to outside the boundary. For both outcomes, boundary effects persist and remain significant as the bandwidth narrows to 30 meters.

4.5 Long-term effects

We now examine the persistence of the difference in property values that emerged by 1864. The first set of estimates (Table 8) looks at characteristics of residents inside and outside of the catchment area boundary using data on socio-economic status compiled by Charles Booth in 1899. By this point in time, all of Soho has piped water, so the water pump catchment areas have truly become irrelevant features of the property market. Nevertheless, we see strong evidence of discontinuities at the border in terms of household poverty level. Specifically, households inside the boundary are 64% more likely to be very poor (the lowest wealth class) and 66% less likely to be middle class (the highest wealth class present in this neighborhood) compared with households just outside the boundary, and the difference is statistically significant at the 5% level.

The second set of estimates (Table 9) considers property assessments from the land tax assessment data in 1936. Two points about the 1936 data are worth noting. First, for reasons that are unclear, data on almost half of properties in the lower third of the parish are missing from the 1936 records that have been digitized, leaving a total of 793 property records from St. James. Figure 10 shows the coverage on a map, where it is clear that specific logs were either never collected, not preserved, or never digitized. Secondly, our sample size is considerably smaller because, by 1936, 45% of properties have been exonerated from taxation.[39] For both reasons our dataset is considerably smaller in this year, although importantly attrition for both reasons is balanced inside and outside the boundary.[40]

As shown in Table 9, property values continue to be significantly higher outside relative to inside the Broad Street pump catchment area in 1936. Once again, the results are robust to considering a relatively narrow and relatively wide bandwidth, and to clustering standard errors by street or including boundary segment fixed effects. In fact, the point estimates, which range from 30% to 37% depending on the specification, suggest that property values have diverged since 1864, although

---

[39] Exoneration at the time of property sale became a law in 1949, and tax was fully abolished in 1963.

[40] The RD estimate of the difference in exoneration rates at the boundary in 1936 yields a point estimate of -0.005 and standard error of 0.074 using the main specification (cubic in distance plus controls and block clusters).

the difference in magnitude across years is not statistically significant.

In Table 10 we conduct the same exercise using data from contemporary St. James (now Soho district). As described previously, since land value assessments are no longer conducted by the state, we use real estate transactions between 1995 and 2013, along with current rental price estimates from an online database. Together, these sources provide 1877 observations on property prices in what used to be St. James parish. Using the same RD specification, we estimate a property value differential inside and outside the boundary of the Broad Street pump catchment area of between 16% and 30%, depending on the data source. Boundary effects on rental prices are also significantly negative, and almost identical to the estimates from 1864.

This pattern is important as well as reassuring. The fact that differences are reflected in actual property transactions and not only data from the Land Tax Assessment assuage concerns over measurement error or potentially biased reporting of rental values in the official records.

4.6 Robustness checks

4.7 John Snow's BSP boundary definition

Here we assess the robustness of the results to the treatment boundary used. Specifically, we compare results using an alternative definition of the BSP catchment area boundary proposed by John Snow in his cholera report shortly after the outbreak (Snow, 1855).

Figure 2a depicts John Snow's original boundary (in blue) and a modification of the boundary that excludes the pump at Little Marlborough St. (in black).[41] For comparison, Figure 2b includes the network Voronoi boundary used in the main analysis. Note the significant overlap between John Snow's modified boundary and the network Voronoi boundary. In fact, for all years, the percentage of houses inside the network Voronoi boundary that are also inside Snow's modified boundary is close to 100%.[42] Following previous literature (e.g., Shiode (2012)), we use the modified version of Snow's boundary in the results below since it excludes the Little Marlborough St. pump which was not being used at the time of the outbreak. Table 11 presents the estimated RD coefficients using

---

[41] Recall that the pump at Little Marlborough St. was not being used at the time of the outbreak (Snow, 1855). We obtain John Snow's modified boundary from Shiode (2012).

[42] Specifically, for the year 1853, 455 out of the 458 houses inside the network Voronoi boundary are also within Snow's modified boundary. The numbers for the remaining years are, 445 out of 448 for 1864, 311 out of 312 for 1936, and 729 out of 734 for the 1995-2013 period.

Snow's boundary as the treatment boundary. Note that the magnitude and statistical significance of the coefficients are similar to the ones obtained using the network Voronoi definition.

4.7.1 Falsification tests

Many other features of the neighborhood may be changing and contributing to cross-block variation in real estate prices that coincide with the emergence of a boundary effect between 1853 and 1864. It is difficult to imagine why changes in amenities would generate a discontinuity within the space of a block or so, least of all one that happens to coincide with the Broad Street pump catchment area. Nonetheless, it is possible that cross-block price differences of the magnitude of our treatment effect estimates are sufficiently common in this setting as to raise concerns about the validity of our interpretation. This is particularly compelling given that our estimates are concentrated in one very small geographic area that may have been undergoing numerous changes at this point in time.

Two points are worth noting. First, our boundary effect estimates are robust to changes in observable characteristics of the neighborhood (e.g. new schools or roads built) that appear on a detailed map of Soho, which we add as controls in the regression. Second, our conditional treatment effect estimates indicate treatment effects at several points on the boundary, making it less likely they are picking up neighborhood characteristics we do not observe that coincide in multiple locations with the emergence of a boundary effect between 1853 and 1864.

Nonetheless, to gauge the likelihood that our treatment effects simply reflect regular fluctuations in prices across space, we examine the distribution of differences and standard deviations in average house prices across neighboring cells of a grid covering the Greater London area, available from a database of contemporary real estate transactions. If sharp differences in prices across neighboring blocks are a common occurrence, the differences in prices we observe across neighboring blocks within the bandwidth of our treatment effect estimates should fall near the center of such distributions.

For this exercise, we geocode 18 years of house sales data from the Land Registry of England (Land Registry, 2014).[43] The data contain information on addresses and sales price for all houses sold in the Greater London area between 1995 and 2013. This yields a total of about 2 million

---

[43] Refer to Figure **??** for a plot of the average real prices by block for the Greater London area. Prices are in 2013 pounds.

transactions.[44] We overlay geocoded addresses on a grid covering the area with each graticule cell having length equal to the optimal bandwidth used to calculate boundary effects on contemporary real estate prices (43 meters). To avoid potential variations in price due to physical barriers that do not characterize our study area, cells that are intersected by a primary road, railway, monorail, river, or canal are excluded from the analysis. For a given cell, adjacent cells are defined as the four cells adjacent to each side of the cell. Refer to Figure **??** for an example of the grid with sample adjacent cells highlighted.

Figures 12a and 12c present the distribution of log price differences and standard deviation of log prices between adjacent cells, respectively. The solid vertical lines give the average difference in log prices between houses inside and outside the BSP boundary (Figure 12a) and the variance in log price within our optimal bandwidth (Figure 12c). Dashed lines indicate key percentiles of the corresponding distributions. Note from Figure 12a that price differences are clustered around zero indicating that prices differ very little for adjacent cells. Similarly, the highly skewed distribution of standard deviations in price suggest that very few adjacent cells exhibit high price variation. This is not the case in our setting. In fact, the standard deviation of prices in Soho is well above the 95th percentile of the distribution while less than 10 percent of adjacent cells exhibit price differences higher than what we estimate as our boundary effect. The evidence remains robust if the distributions are limited to cells with average prices within the range observed in Soho (Figures 12b and 12d)

In a related exercise, we replicate the RD design using the catchment areas of pumps that were not the source of the cholera outbreak. The purpose of this exercise is to assess the validity of the BSP boundary results by comparing property values across the boundaries of relatively unaffected catchment areas. We refer to the boundaries in this exercise as *false treatment boundaries*.

The choice of a false treatment boundary is determined by data availability near the boundary. Figure 10 presents four false boundaries highlighted in red along with the full sample and the estimation sample around the boundary highlighted in green. For reference, Figure 10 shows all pumps and their respective catchment areas within Soho. When a false pump is adjacent to BSP, we exclude observations falling inside the BSP area. We choose bandwidths around the boundaries

---

[44] For the analysis presented in this section, we use information on about 1.2 million transactions as the geocoding process is not complete yet. Results will be updated once the process ends.

following Imbens and Kalyanaraman (2012).

Panels A through D in Table 12 present the RD coefficients from the estimation of Equation (1) using each of the four boundaries in Figure 10 as the treatment boundaries. We use false boundaries 1, 2, and 3 in the pre-outbreak (1853) and post-outbreak (1864) periods (Columns (1)-(6)), and false boundaries 2 and 4 for the exercise using current property values (Columns (7) and (8)). Because of data availability, we cannot perform any falsification tests for the 1936 data.

Column (1) shows the results for the pre-outbreak period (1853). Similar to the pre-outbreak results at the BSP boundary, there is no evidence of a pre-outbreak “pump effect”. This is an important result considering that the presence of such effect could confound the cholera effect found at the BSP boundary. Columns (2) and (3) shows that no significant changes in deaths occur at the boundaries of unaffected pumps. This is expected considering that BSP is the contaminated pump.[45] Columns (4) and (5) give RD estimates for rental prices in 1864 and change in rental price between 1864 and 1853, respectively. For all three cases, the difference in property prices across the boundaries tested are not statistically significant. For boundaries 1 and 3, the magnitude of the difference in rental prices (around 5%) is less than half the magnitude observed at the BSP boundary (around 13%). Column (6) shows that the likelihood of a change in residency from 1853 to 1864, although relatively high in the case of false boundaries 1 and 2, is not statistically significant. In the case of boundary 3, this likelihood is almost zero. Lastly, Column (7) reports differences in sales price and Column (8) shows *Zoopla* house value estimates properties near boundaries 2 and 4. Unlike the results observed at the BSP boundary, house prices and estimated values vary smoothly across the false boundaries. In addition, the estimated RD coefficients are statistically insignificant in all cases.

4.8 Possible Interpretations

Before turning to our interpretation of the mechanism underlying the change in real estate prices at the border, outlined in the introduction, we discuss the validity of two alternative interpretations for the patterns that emerged.

---

[45] Note that, although the difference in the likelihood of a house with at least one death at boundary 3 is marginally significant, the magnitude of the coefficient is less than half the magnitude observed across the BSP boundary.

4.8.1 Stigma of disease

It is possible that houses hit by cholera suffered a loss in value not because of the general disutility of living among newly impoverished neighbors, but because the deaths from cholera carried a particular stigma that reduced renters' willingness to live inside the catchment area. While it is impossible to rule out any role of stigma, two points are worth noting. First, we observe no difference in property prices inside the BSP area between houses hit and houses not hit by cholera, conditional on the fraction of other houses on the block that experienced a cholera death (Table 13). Hence, there is no evidence of property-specific stigma effects. Block-level stigma effects - the idea that properties on a block that was heavily hit by cholera fall in value because of prospective renters' fear of inhabiting that block - are virtually impossible to disentangle from other sources of disutility prospective renters might experience from living among the poor. However, it is arguable that any income shock that leads to poverty will carry with it a certain amount of stigma, so in that sense the distinction may be irrelevant to the external validity of our results.

Second, the RD design once again works against a pure stigma story in that, for stigma to produce strong boundary effects, prospective renters would need to have a clear idea of the exact location (within 60 meters) at which disease rates rapidly fell. While newspaper accounts frequently discussed the cholera epidemic in the context of Soho and Broad Street pump, and even specific streets within the neighborhood that were heavily hit, it was unlikely there were references to specific blocks. Hence, it is hard to imagine that stigma would change sharply along an invisible boundary unless it is the stigma of poverty more generally, which would be visible to prospective renters.

4.8.2 Devaluation of water infrastructure

Is it possible that the price differences that emerge as a result of the cholera epidemic are simply the market reaction to updated beliefs about the quality of the local water source or disease environment more generally (e.g. bad air quality, or "miasma" that many people at the time believed to be the source of the epidemic)? The main advantage of a regression discontinuity approach to estimating disease effects is to rule out any such story. That is, even if the water pump was deemed completely unusable, such that all residents of St. James relied on the other twelve pumps post-epidemic, we would not expect to see a jump in real estate prices at the boundary of the catchment area. Real estate prices would decrease smoothly with distance to pump, which should be absorbed by the RD polynomial.

Furthermore, the fact that boundary effects persist over 150 years indicates that they do not simply reflect beliefs about the quality of the local water source or disease environment. By the end of the 19th century, piped water had replaced well water in Soho, so property prices could not possibly reflect updated beliefs about well quality. Similarly, beliefs about other sources of disease transmission attributed to specific locations would have disappeared within 80 years.

5 Theoretical analysis

In this section we investigate the model of a rental market in which an unexpected and localized negative income shock hits some of the tenants on particular blocks. We are particularly interested in characterizing conditions for such a shock to be able to permanently change the composition of renters.

5.1 Baseline model

For simplicity, in the baseline model we assume that there are two types of tenants, rich (r) and poor (p). We consider the problem of a single profit-maximizing owner of a block with $n \geq 2$ apartments, after some tenants on a block of previously rich tenants are hit by a disease shock and became poor. Later we will extend the analysis to the case when there are multiple landlords on the block. We consider a discrete time model with time periods $t = 0, 1, 2, ...$, where $t = 0$ is normalized to be the first instance after the shock that a rental agreement pertaining to one of the apartments on the block is renegotiated. We assume that, after the shock, $x \in \{0, ..., n-1\}$ of the current tenants are poor and the remaining $n-1-x$ are rich. At every subsequent period, there is a probability $q \in (0, 1]$ that a rental agreement on the block (uniformly randomly selected) is renegotiated. These renegotiation opportunities arise partly because existing contracts with tenants expire at idiosyncratic times, but possibly also because a tenant finds an outside option that makes her better off than remaining in the current apartment with the current rent. For simplicity we model all these events through a single time-independent stochastic process. A key features of this set-up is that the composition of the block can only change gradually.[46]

We assume that tenants have additively separable utility functions in housing and money. The per period utility a tenant obtains when living on the block depends on the composition of the block. We assume that poor residents on the same block exert a negative externality on their neighbors.[47] In particular, the utility a type $s \in \{p, r\}$ tenant obtains when paying rent r is $-r - c_k^s$, where k is the number of poor other tenants in the block. Hence $c_0^s - c_k^s$ can be interpreted as the premium that a type s tenant is willing to pay not to have any poor other tenants in the block, relative to having k of them. As a simplifying assumption, for most of the analysis we assume that $c_k^p = 0$ for every $k \in \{0, ..., n-1\}$ – that is, poor tenants' willingness to pay to reduce the number of poor neighbors is zero – but show in an extension that our results extend to allowing poor tenants to have positive willingness to pay for avoiding poor neighbors as long as it is less than that of rich tenants. We assume that $c_0^r \geq 0$, and that c_k^r is strictly increasing in k. We allow for the possibility of $c_0^r > 0$ because, even if there is no current poor tenant on her block, poor residents of neighboring blocks might exert negative externality on a rich tenant. If one assumes there are no negative externalities across blocks then it is natural to set $c_0^r = 0$. The outside option of a type s tenant is $-W^s$ per period, which can be interpreted as living at another location where rent is W^s and there are no poor neighbors. To make the landlord's problem nontrivial, we assume that $W^r - c_0^r > W^p$.

An important assumption that we maintain is that the area of impact of the negative shock, and in particular the size of the block, is small relative to the whole economy and therefore whatever strategy the landlord follows has no influence on rent levels outside the block.

The landlord maximizes the expected present value of current and future rents, taking into account the fact that the composition of tenants on the block influences the amount of rent a rich renter is willing to pay. All agents are fully forward-looking and discount future payoffs by a

---

[46] While it is advantageous for a landlord to synchronize the timing of renegotiations across apartments on the block, that is unlikely to be possible given idiosyncratic turnover of tenants coupled with the discrete nature of tenancy agreements at the time, which took the form of 1-, 3-, 5- and 10-year contracts, and were governed by different regulation of terms (such as how many months of non-payment before property could be confiscated from the tenant as compensation, etc.).

[47] Further motivation is provided in Guerrieri et al. (2013), who impose a similar assumption. We note that as the levels of utilities are free to move around in our model, we can equivalently think about assuming that rich neighbors exert positive externalities on tenants, instead of poor neighbors exerting negative externalities.

factor δ. In the baseline model we allow the landlord to perfectly screen potential new tenants and essentially choose the type of tenant.[48] The amount of rent the tenant is willing to pay depends on the tenant's type, and in case of a rich tenant, on both the current composition of tenant types on the block and the expected composition in the future (which depends on the landlord's future expected choices). The landlord is indifferent between renegotiating the contract with an existing tenant or acquiring a new tenant of the same type as the current tenant, since the maximum amount of rent he can get is the same. Similarly, a tenant is indifferent between renegotiating the contract or moving out if the rent offered makes her exactly as well off as her outside option. We assume that in such cases parties choose to renegotiate the rental contract (which can be motivated by small transactions costs associated with moving on the tenant's side, and with acquiring a new tenant on the landlord's side), hence a tenant only moves out for non-exogenous reasons if she is replaced with a different type of tenant.

Technically, the landlord's problem is not a simple one-person decision problem, because the rent he can charge at different negotiations depends on tenants' expectations of the landlord's future actions. We assume that the landlord can choose a strategy at the beginning of the game, and that tenants have correct expectations of future actions dictated by this strategy. We also show that the optimal strategy for the landlord is sequentially rational, so it is never in his interest to depart from it.

5.2 Main predictions

The first prediction we obtain from the baseline model is that the landlord's optimal strategy is either to (i) retain all rich tenants and over time fill all new vacancies with new rich tenants, or (ii) retain all poor tenants and over time fill all new vacancies with new poor tenants (more precisely, at least one of these strategies is always among the optimal strategies). We will refer to these strategies as "always rich" and "always poor." The intuition is that, if at a certain state it is optimal to acquire a rich (respectively, poor) tenant, then it remains optimal to do so in future times when the ratio of rich (poor) tenants is higher. A landlord following the always rich strategy finds sticking to the strategy more and more profitable over time, since, as the block transitions to

---

[48] In an alternative version of the model below we show how the main results extend to a setting in which the landlord cannot directly discriminate between tenant types, only through posted prices.

rich, he can charge rich tenants higher and higher rents. Similarly, a landlord following the always poor strategy finds it less and less profitable to deviate and go after a rich tenant, since over time he has to offer more discount for a longer time to rich types in order to attract them. While the intuition is simple, the precise statement and proof of the result is technical, and so is relegated to the Appendix.

Whether the landlord should choose the always rich versus the always poor strategy depends on the number of current poor tenants on the block, x, at the time of the first vacancy. If the block is hit by a severe enough negative income shock so that x is larger than a critical threshold, it can be too costly for the landlord to start acquiring rich tenants and build back an all-rich block. Instead, it becomes optimal to let the remaining rich tenants move out and let the block become poor. Notice that, in case of an all poor strategy, rents from all future new tenants and from all current tenants staying after the first renegotiation is independent of x, and equal to W^p. However, expected payoffs of the landlord when following an always rich strategy strictly decrease in x.

The next result characterizes the critical threshold determining the landlord's optimal strategy.

Proposition 1: The landlord prefers the always rich strategy to the always poor strategy iff:

$$W^r - W^p > \sum_{i=0}^{x} \frac{(1-\delta)\frac{x!}{i!}(x+1-i)(\delta\frac{q}{n})^{x-i}(1-\delta(1-\frac{q}{n}))}{\prod\limits_{m=i}^{x+1}(1-\delta(1-q\frac{m}{n}))} c_i^r \tag{3}$$

Note that an increase in W^r relative to W^p increases the left hand side of (3), and so increases the threshold x^* at which the landlord switches to the always poor strategy. The right hand side of (3) is increasing in each c_i^r ($i \in \{0, ..., x\}$), hence an increase in any of these cost parameters decreases the threshold. This in particular holds for c_0^r, which means that in case there are across block externalities, an increase in the number of poor tenants in neighboring blocks makes it less likely, ceteris paribus, that the landlord chooses the always rich strategy. The comparative statics in δ are more complicated, but as $\delta \to 1$ the right hand side of (3) converges to c_0^r, hence, given our assumption of $W^r - W^p > c_0^r$, a very patient landlord chooses the always rich strategy for any x.

Also note that the landlord choosing the always rich strategy implies that the initial rich tenants stay on the block, while the initial poor tenants gradually move out. The landlord choosing the always poor strategy implies the opposite: initial poor tenants stay on the block, while initial rich tenants move out.[49] Combining this with Proposition (1) establishes that an increase in the degree to which the block is affected by the negative shock, as summarized by x, increases the rate at which rich tenants move out relative to the rate at which poor tenants move out.

While the derivation of condition (3) for general n is relegated to the Appendix, here we demonstrate the derivation for the case of $n = 2$, when the existing tenant at the start of the game is a poor type.[50] We highlight this case because, for tractability, most of the extensions of the baseline model provided below are in the context of $n = 2$.

Consider a rich tenant, who pays rent r per period. If she realizes her outside option, she gets $V(out) = -\frac{W^r}{1-\delta}$. It is easy to see that the rent that makes the tenant indifferent between renting versus the outside option is $r = W^r - c_0^r$, which is the rent the landlord can negotiate with rich types if their neighbor is rich. Now let $V(poor)$ and $V(rich)$ denote the expected continuation utility of a rich tenant renting an apartment for a general fixed r, given a poor and a rich neighbor, assuming that the landlord is following the always rich strategy. Next period three situations are possible: no change, the neighbor's rental contract is renegotiated, or the tenant's contract is renegotiated:

$$\begin{aligned} V(rich) &= -(r + c_0^r) + \delta[(1-q)V(rich) + \frac{q}{2}V(rich) + \frac{q}{2}V(out)] \\ V(rich) &= \frac{-(r + c_0^r) + \delta\frac{q}{2}V(out)}{1 - \delta(1 - \frac{q}{2})} = \frac{-(r + c_0^r) - \frac{q}{2}\frac{\delta}{1-\delta}W^r}{1 - \delta(1 - \frac{q}{2})} \end{aligned} \tag{4}$$

and

$$V(poor) = -(r + c_1^r) + \delta[(1-q)V(poor) + \frac{q}{2}V(rich) + \frac{q}{2}V(out)].$$

A profit-maximizing landlord chooses rent r^* such that $V(poor) = V(out)$. Hence:

$$r^* = W^r - \frac{\frac{1}{2}\delta q}{1 - \delta + \delta q}c_0^r - \frac{1 - \delta + \frac{1}{2}\delta q}{1 - \delta + \delta q}c_1^r$$

To summarize, a landlord following an always rich strategy at the beginning of the game acquires

---

[49] If we made the model more realistic by also allowing for tenants moving out for exogenous reasons then over time both rich and poor tenants leave the block, but if the landlord chooses the always rich strategy, initial poor tenants in expectation leave earlier than initial rich tenants. Similarly, in case the landlord chooses the always poor strategy, initial rich tenants in expectation leave earlier than initial poor tenants.

[50] If $n = 2$ and the existing resident at the beginning of the game is a rich type then independently of other parameters, the landlord can achieve his maximum possible payoff in the game by the always rich strategy,

a rich tenant, and negotiates a rent of r^*. Then at the first renegotiation opportunity with the other tenant, he lets the tenant leave and acquires a new rich tenant, with a rent of $W^r - c_0^r$. This rent prevails in all future negotiations. Given this, the landlord's payoff when following the always rich strategy, net of the exogenously given rents paid by the initial poor renter, is:

$$U_{rich} = \frac{(1+\frac{\delta q}{1-\delta})W^r - \frac{\delta q}{1-\delta}c_0^r - c_1^r}{1-\delta(1-\frac{q}{2})}$$

Using the fact that, in case of an always poor strategy at the beginning of the game and at every future negotiation a rent of W^p is agreed upon, the expected payoff from following an always rich strategy yields a higher payoff than following an always poor strategy iff:

$$W^r - W^p > \frac{\delta q}{1-\delta+\delta q}c_0^r + \frac{1-\delta}{1-\delta+\delta q}c_1^r.$$

5.3 Extensions of the model

Investments/maintenance

Suppose the landlord in every period has to make an additional choice of making either high investment/maintenance (H) into the block, or low investment/maintenance (L). The cost of L is normalized to 0. The cost of H per period is $k > 0$. For simplicity, assume that poor tenants do not care about the level of investment, but rich tenants in each period suffer a disutility of d when the previous investment decision was L.[51]

Assume that a cost-to-disutility ratio $\frac{k}{d}$ is low enough that in case of the "always rich" strategy it is profitable to always choose H. In the Supplementary Appendix we show that this is equivalent to assuming that:

$$\frac{k}{d} \leq \frac{\delta\frac{q}{n}[n - x\delta + \frac{x(n+1)}{n}\delta q]}{1-\delta+\frac{x+1}{n}\delta q}.$$

If this condition holds then in the case of always rich strategy is chosen by the principal, it is always accompanied by investment level H, and all rents stay the same as in the baseline model. However the owner has extra losses from the costs of H, resulting in the expected payoff from

---

[51] The results below readily generalize to the case when rich tenants have higher willingness to pay for H vs L investment than poor tenants.

following an always rich strategy decreasing to:

$$\frac{(1-\delta+\delta q)}{(1-\delta)(1-\delta(1-\frac{q}{n}))}W^r - \sum_{i=0}^{x} b_{ix}c_i^r - \frac{k}{1-\delta}.$$

This implies that the model with investments is equivalent to the baseline model with $(W^r)' = W^r - \frac{1-\delta+\delta\frac{q}{n}}{1-\delta+\delta q}k$ instead of W^r. Therefore, the qualitative conclusions of the model are the same as before, but an increase in the cost of H investment make choosing the always rich strategy less profitable, hence making it more likely that the landlord's optimal strategy is always poor.

Poor types also willing to pay premium for rich neighbors

As long as we assume that $c_k^r - c_{k-1}^r > c_k^p - c_{k-1}^p$ for every $k \in \{1, ..., n-1\}$, that is the marginal willingness to pay to reduce the number of poor neighbors is always higher for rich types than for poor types, the result that either the always poor or the always rich strategy is optimal continues to hold. In the Supplementary Appendix we derive the conditions in this extended model for the optimality of the always rich versus the always poor strategy. Fixing all other parameters, increasing any of the cost parameters c_k^p for $k \in \{x, ..., n-1\}$ decreases the payoffs from the always poor strategy, while not affecting the payoffs from the always rich strategy. In the case of $n = 2$, the condition for always rich being an optimal strategy is the following simple modification of the original condition:

$$W^r - W^p > \frac{\delta q}{1-\delta+\delta q}c_0^r + \frac{1-\delta}{1-\delta+\delta q}c_1^r - c_1^p.$$

Multiple owners

If not all apartments on the block are owned by the same owner, there are additional coordination issues arising among owners, as well as a free rider problem (it is better if another owner starts changing the composition of the block at the expense of current losses) and multiplicity of equilibria. The latter might result in the block converging to being all poor even when owners are very patient. The fact that tenants receive asynchronous opportunities to move out can still imply that in all equilibrium ultimately the block converges back to being all rich,[52] but this requires a more demanding condition than the one implying that the all rich strategy is optimal for a single

---

[52] On how asynchronicity of moves can solve coordination problems, see for example Lagunoff and Matsui (1997), Takahashi (2005), Dutta (2012), Calcagno et al. (2014) and Ambrus and Ishii (2015).

owner. In short, multiple owners make it more likely that after a concentrated negative income shock the block converges to be all poor, and less likely that it converges back to be all rich. We demonstrate this in the case when there are two apartments, owned by two different owners. We restrict attention to Markov perfect equilibria of the game between the landlords, which for brevity we just refer to as equilibria.

First, note that a necessary condition for there to exist an equilibrium with two owners such that an owner is willing to acquire a rich tenant when the current tenant in the other apartment is poor is that it is profitable to do so assuming that this triggers the other owner to change his tenant to a rich type, at the first possible opportunity in the future. This also turns out to be a sufficient condition for the existence of an equilibrium in which the block converges to all rich, in case $x = 1$.

The maximal rent a rich type is willing to accept given the above profile is:

$$r^* = W^r - \frac{\frac{1}{2}\delta q}{1 - \delta + \delta q} c_0^r - \frac{1 - \delta + \frac{1}{2}\delta q}{1 - \delta + \delta q} c_1^r.$$

Given this rent and the above strategy profile, the landlord's expected payoff is:

$$U_{rich} = \frac{W^r}{1 - \delta} - \frac{\delta \frac{q}{2}}{(1 - \delta)(1 - \delta(1 - \frac{q}{2}))} c_0^r - \frac{1}{1 - \delta(1 - \frac{q}{2})} c_1^r$$

If instead the landlord always hires poor tenants, then his utility is $U_{poor} = \frac{W^p}{1-\delta}$. The apartment owner prefers the always rich strategy when:

$$W^r - W^p > \frac{\delta \frac{q}{2}}{1 - \delta(1 - \frac{q}{2})} c_0^r + \frac{1 - \delta}{1 - \delta(1 - \frac{q}{2})} c_1^r. \tag{5}$$

Note that this condition is stricter than the condition for a monopolist landlord's optimal strategy being always rich. Hence multiple landlords on the block make it more likely that a block hit by a negative income shock transitions to poor, even if the best equilibrium is played by the landlords.

Moreover, even when condition (5) holds, there might be another equilibrium, caused by the co-ordination problem between the two landlords, in which both landlords follow the all poor strategy.

In the Supplementary Appendix we show that such equilibrium can be ruled out iff:

$$W^r - W^p > \frac{1-\delta+\frac{1}{2}\delta q}{1-\delta+\delta q}c_0^r + \frac{\frac{1}{2}\delta q}{1-\delta+\delta q}c_1^r. \tag{6}$$

No price discrimination

In the baseline model we assumed that a landlord can perfectly discriminate between rich and poor types, effectively choosing which type of tenant he wants to fill a vacancy. Here we focus on the case of $n = 2$, and show that even if such discrimination is not possible, and the landlord can only choose a posted rent for a vacancy, having to accept any tenant willing to pay the posted rent, the qualitative conclusions of the model remain unchanged. Moreover, it becomes *more* likely that the landlord chooses the always poor strategy.

If the apartment owner cannot discriminate against poor applicants, and the maximal rent a rich tenant is willing to accept is less than what a poor tenant is willing to pay (because of the current high number of poor tenants) then a posted price equal to the maximal willingness to pay of the rich types attracts both types of tenants. In such cases we assume that the probability that the tenant accepting the offer is a rich type is $\pi \in (0, 1)$. Let r^* be the maximal rent that a rich type is willing to pay when the current other tenant is poor, but at the first possible renegotiation opportunity she is expected to be switched to a rich tenant. Assume $r^* < W^p$, so hiring a poor tenant has short-term benefits for the landlord.

In the Supplementary Appendix we show that in this modified environment the always rich strategy yields a higher payoff for the landlord than the always poor strategy iff:

$$W^r - W^p > \frac{\delta\frac{q}{2}[(1-\delta)(1+\pi)+2\delta q\pi]}{(1-\delta+\delta q)(1-\delta+\delta q\pi)}c_0^r + \frac{(1-\delta)(1-\delta+\delta\frac{q}{2}(1+\pi))}{(1-\delta+\delta q)(1-\delta+\delta q\pi)}c_1^r.$$

This condition is stricter than the condition for the always rich strategy being more profitable than the always poor strategy in the baseline model, hence inability of the landlord to price discriminate increases the likelihood that a block hit by a negative income shock transitions to be all poor.

Gentrification

Differences between two blocks in type composition, created by random locally correlated shocks, can prevail even after a general increase in demand for housing in the district (comprising both blocks) that shifts the type distribution in both blocks towards wealthier tenants. Such a trend characterizes Soho over the last two decades, during which time average sales prices have increased by 139%. Meanwhile, our empirical results indicate that the wedge in rental prices remains even as the district has gentrified.

To demonstrate how this is possible in the context of our model, we extend the baseline model to include four types of prospective renters: poor, middle-class, rich and very rich. Their outside options are correspondingly $-W^p$, $-W^m$, $-W^r$, $-W^v$ per period, where $W^p < W^m < W^r < W^v$. Let c_i^m be the cost to a middle-class tenant of having i poor neighbors, and assume it is increasing in i. Also assume that $W^r - c_{n-1}^m < W^p$, but $W^r - c_0^m > W^p$. Let $c_{i,j}^r$ be the cost for a rich tenant of having i poor neighbors and j middle-class neighbors, and let $c_{i,j}^r$ be increasing in both i and j. Furthermore, assume that if $i + j = i' + j'$ and $i > i'$ then $c_{i,j}^r > c_{i',j'}^r$. Let $c_{i,j,k}^v$ be the cost for a very rich tenant imposed by having i poor neighbors, j middle-class neighbors and k rich neighbors, and let $c_{i,j,k}^v$ be increasing in i , j and k. Assume also that if $i + j = i' + j''$ and $i > i'$, then $c_{i,j,k}^v > c_{i',j',k}^v$ and if $j + k = j' + k'$ and $j > j'$, then $c_{i,j,k}^v > c_{i,j',k'}^v$. Lastly, assume that $W^p < W^m - c_0^m < W^r - c_{0,0}^r < W^v - c_{0,0,0}^v$. Intuitively, these assumptions imply that all types are willing to pay a premium to avoid having neighbors of lower type, and higher types have a higher willingness to pay.

In the Supplementary Appendix we show that there is a parameter range for which originally both an all poor and an all rich block are stable, and after the increase in the attractiveness of the district, the poor block transitions to a middle-class one, while the rich block transforms to a very rich one. Therefore, it can be the case that, if two originally rich blocks are hit by a negative income shock differentially, one converges back to rich and one slides down to being poor, and there remains a difference between these blocks even if later the composition of types transitions upwards in both of the blocks, due to an exogenous increase in the attractiveness of the blocks.

5.4 Back of the Envelope Calculations

The results above show that it is theoretically possible that, when a negative income shock hits many of the current tenants, a profit-maximizing landlord chooses to let the block transition from rich to poor, despite the fact that an all rich block would make him better off in the long run. However, to lend credibility to our interpretation of the empirical results, it is important to establish that this can be the case for realistic parameter values, for example without requiring the landlord's level of impatience be implausibly high. Here we provide some back of the envelope calculations to verify that the all poor strategy can indeed be optimal for plausible parameter values.

Below we assume a linear disutility function for rich types from poor neighbors: $c_x^r = x \times y$, where y is the incremental disutility of an extra poor neighbor, in a block of 40 apartments (the average block size in the parish). Based on Allen (2005, 2009), we set the returns to capital around the time of the cholera epidemic to 20%, hence the relevant discount factor for a landlord to $\frac{1}{1.2}$. Our estimate of the price difference between the rent a landlord can get in an all rich block versus in an all poor block ten years after the epidemic (by which time transitions finished with very high probability) is 26%, motivating us to set $W^p = 1$ and $W^r = 1.26$. We have no data on the differential investment and maintenance costs for blocks targeting rich tenants, and therefore make alternative calculations using d ranging from 0 to 0.1 (the latter decreasing the revenue difference for the landlord to 16%). We set the time between periods to be a week, and set q such that contracts on average get renegotiated in every two years.

Suppose the critical threshold for a block transitioning to poor is 25%, that is ten initial poor households in a block of 40 apartments. For $d = 0$, the incremental disutility in the linear disutility functional specification that makes a landlord indifferent between the always poor and the always rich strategy at $x = 10$ is $y = 0.168$. With this disutility parameter, the initial rent that the landlord would have to offer to retain a poor tenant is 0.29, implying a 71% discount relative to W^p. For $d = 0.1$, the disutility parameter making the landlord indifferent decreases to $y = 0.104$, increasing the initial rent offered to rich tenants to 0.56 (a 44% discount).[53]

We think the above levels of disutilities for additional poor neighbors, and relatedly, the amount of discount a landlord initially has to offer to retain rich tenants, are not unreasonably high.

---

[53] Please contact the authors for the Mathematica files with the computations.

Furthermore, recall from the previous analysis that neighboring poor blocks (which are more likely for a block hit hard by the negative income shock) increase the relative attractiveness of the all poor strategy, leading to even smaller levels of disutility from poor neighbors required to rationalize such a strategy. Similarly, multiple owners within the block make it more likely that the owners choose the all poor strategy, again reducing the level of disutility from poor neighbors that is required for the block to converge to all poor. Thus, both of the above aspects decrease the incremental disutility parameter implying a threshold of 25% for the block transitioning to poor (and also the initial discount required to retain a rich tenant at the threshold ratio of initial poor households). We also did computations with a strictly concave disutility function (the very first poor neighbor imoses the highest incremental disutility, the second one the next highest, etc.), and for the same parameter values the initial discounts that a landlord has to offer to retain a rich tenant at the threshold are lower than in the linear disutility specification.

The above computations are based on a relatively low discount factor of $\frac{1}{1.2}$. Over time, the returns to capital decreased to levels significantly lower than the 20% Allen (2005) estimates for around 1860. This raises the question of whether blocks that transitioned to poor should be expected to transition back to rich, given the above parameter values of the model, once the discount factor increases to recent levels. The answer is no. Once transitioning finished and all current tenants are poor, in all of the cases considered above the estimated disutility parameters imply that always poor remains the optimal strategy for the landlord, even when interest rates decrease to 5-10%. Moreover, if we assume that the high return to capital prevailed throughout the period of heavy industrialization, until around 1915,[54] then even when assuming that landlords in 1860 foresaw the drop of capital returns after this period to 5-10%, the threshold level of x that makes the landlord indifferent between the always poor and the always rich strategy remains essentially unchanged, as 55 years is far enough in the future that with the above interest rates it only has a miniscule impact on the the landlords' expected payoffs.

The above calculations show that the outcome that blocks transition to permanently all poor once a significant fraction of them become poor is not just a theoretical possibility, but that it can happen with reasonable parameter values.

---

[54] This is exactly the period during which the number of rail passengers grew exponentially in Great Britain. See Association of Train Operating Companies (2008)

6 Other contributing factors

Aside from the mechanism we highlight in this paper, several other factors are likely to have contributed to the persistence of income differences at the boundary of the cholera epidemic. First, the optimal property investment path of a landlord should depend on the landlord's expectation that the block remains poor. Hence, for houses on blocks above the threshold level of impoverishment, the epidemic should reduce incentives to invest in property, making it even more likely that the block gets stuck in a poor equilibrium over the long run.

Second, demographic trends could play a similar role if renters derive additional disutility from living among ethnic minorities such as Irish and Jewish immigrants, who moved into Soho in large numbers in the late nineteenth and early twentieth centuries. If immigrants sort onto slightly lower-priced blocks, this will further encourage low-rent blocks to remain so over time since it further lowers the willingness of the rich to live in a poor neighborhood, and hence the discount a landlord would need to offer them.

Another potential contributing factor is the license procedure for sex establishments, of which there have historically been many located in the Soho district. Essentially the city council has full jurisdiction over which establishments are granted licenses, which could lead to a sorting of SEV onto impoverished neighborhood blocks. Assuming such establishments generate negative externalities on residents of the same block, the segregation of SEVs could contribute to persistent differences in the sorting of individuals across neighborhood blocks.

Finally, a major factor contributing to the persistence of residential patterns in the twentieth century are tenancy laws that were in effect between 1915 and 1985, which gave existing tenants extremely strong occupancy rights and rent control. The Increase of Rent and Mortgage Interest (War Restrictions) Act of 1915 restricted the right of landlords to eject their tenants and prevented them from raising the rent except for limited purposes. Before the 1915 Act, the relationship between landlord and tenant had been purely contractual; at the expiration or termination of the contract, the landlord could recover possession. Various rent control laws went into effect until the Housing Act of 1988, which almost fully deregulated the rental market.

7 Conclusion

Our findings provide novel evidence that idiosyncratic shocks to individuals can have a permanent effect on the spatial distribution of poverty within a city, even in a thick rental market with few frictions in which only renters (rather than owners) are shocked. More broadly, they imply the existence of a simple channel through which we may observe persistence of historic events in any setting - the resorting of individuals can put a neighborhood onto a different growth trajectory even when its infrastructure is untouched.

As a result, one potential cost of spatially correlated shocks is the resulting misallocation of land if entire blocks house lower income residents than is optimal according to their intrinsic value. Such a possibility provides rationale for third-party interventions such as “urban renewal” projects or other attempts to upgrade poor neighborhoods located on intrinsically valuable property. On the other hand, the sorting process also implies a form of insurance to those who experience disease or other income shocks that are spatially correlated: the more that their network is hit, the less likely they are to be priced out of their neighborhood. The smaller scale the spatial variability, the more valuable it is for the newly poor to remain on previously high rent land.

References

Ambrus, A. and Y. Ishii (2015). On asynchronicity of moves and coordination. *Economic Research Initiatives at Duke (ERID) Working Paper No. 185*.

Ancestry.com (2011). *London, England, Land Tax Records, 1692-1932 [database on-line]*. Provo, UT, USA: Ancestry.com Operations, Inc.

Association of Train Operating Companies (2008). The billion passenger railway. lessons from the past: Prospects for the future. `www.atoc.org/npsA67D_tmp.pdf`. [Online; accessed 7-September-2015].

Beeton, I. (1861). *Mrs. Beeton's Book of Household Management*. S.O. Beeton.

Black, D. A., J. Galdo, and J. A. Smith (2007). Evaluating the regression discontinuity design using experimental data. *Working Paper, University of Michigan*.

Black, S. (1999). Do better schools matter? parental valuation of elementary education. *Quarterly Journal of Economics 114*(2), 577–599.

Bleakley, H. and J. Lin (2012). Portage and path dependence. *Quarterly Journal of Economics 127*(2), 587–644.

Booth, C. (1902). *Life and Labour of the People in London*, Volume 1. London: Macmillan.

Calcagno, R., T. Sugaya, Y. Kamada, and S. Lovo (2014). Asynchronicity and coordination in common and opposing interest games. *Theoretical Economics 9*(2), 409434.

Cholera Inquiry Committee (1855). *Report on the Cholera Outbreak in the Parish of St. James, Westminster, during the Autumn of 1854*. London: J. Churchill.

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

Cooper, E. (1854). Report to the Metropolitan Commission of Sewers on the house-drainage in St. James, Westminster during the recent cholera outbreak. *London: Metropolitan Commission of Sewers*.

Dell, M. (2010). The persistent effects of Peru's mining mita. *Econometrica 78*(6), 18631903.

Dowell, S. (1965). *A History of Taxation and Taxes in England* (3 ed.). Frank Cass.

Dutta, P. K. (2012). Coordination need not be a problem. *Games and Economic Behavior 76*(2), 519–534.

Erwig, M. (2000). The graph Voronoi diagram with applications. *Networks 36*, 156–163.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and its Applications.* CRC Press.

Farrell, J. and G. Saloner (1985). Standardization, compatibility, and innovation. *Rand Journal of Economics 16*(1), 70–83.

Guerrieri, V., D. Hartley, and E. Hurst (2013). Endogenous gentrification and housing price dynamics. *Journal of Public Economics 100*, 45–60.

Hanlon, W. (2014). Temporary shocks and persistent effects in the urban system: Evidence from British cities after the U.S. civil war. *NBER Working Paper No. w20471, UCLA*.

Holmes, T. (1998). The effect of state policies on the location of manufacturing: Evidence from state borders. *Journal of Political Economy 106*(4), 667–705.

Hornbeck, R. and D. Keniston (2014). Creative destruction: Barriers to urban growth and the great Boston fire of 1872. *NBER Working Paper No. 20467*.

Imbens, G. and K. Kalyanaraman (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies 79*(3), 933–959.

Imbens, G. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615–635.

Imbens, G. and T. Zajonc (2011). Regression discontinuity design with multiple forcing variables. *Working Paper*.

Johnson, S. (2007). *The Ghost Map: The Story of London's Most Terrifying Epidemic–and How It Changed Science, Cities, and the Modern World.* New York: Riverhead Books.

Kane, T., S. Riegg, and D. Staiger (2006). School quality, neighborhoods, and housing prices. *American Law and Economic Review 8*(2), 183–212.

Keele, L. and R. Titiunik (2013). Geographic boundaries as regression discontinuities. *Working Paper*.

Lagunoff, R. and A. Matsui (1997). Asynchronous choice in repeated coordination games. *Econometrica 65*(6), 1467–1477.

Lalive, R. (2008). How do extended benefits affect unemployment duration? a regression discontinuity approach. *Journal of Econometrics 142*(2), 785–806.

Land Registry (2014). Price paid data. `http://bit.ly/1HNQAiA`. [Online; accessed 19-December-2014].

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. house elections. *Journal of Econometrics 142*(2), 675–697.

Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature 48*(2), 281–355.

Maskin, E. and J. Tirole (1988a). A theory of dynamic oligopoly i: Overview and quantity competition with large fixed costs. *Econometrica 56*(3), 549–569.

Maskin, E. and J. Tirole (1988b). A theory of dynamic oligopoly ii: Price competition, kinked demand curves, and education cycles. *Econometrica 56*(3).

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698714.

Mobius, M. (2000). The formation of ghettos as a local interaction phenomenon. *mimeo MIT*.

O'day, R. and D. Englander (1993). *Mr Charles Booth's Inquiry: Life and Labour of the People in London Reconsidered*. London: Hambledon Press.

Okabe, A., B. Boots, K. Sugihara, and S. N. Chiu (2000). *Spatial tessellations: concepts and applications of Voronoi diagrams* (2 ed.). Wiley.

Ordnance Survey (1951). *Ordnance Survey: National Grid Maps, 1940's-1960's. 1:1,250/1:2,500.* Edinburgh, UK: National Library of Scotland.

Pancs, R. and N. J. Vriend (2007). Schelling's spatial proximity model of segregation revisited. *Journal of Public Economics 91*(1), 1–24.

Reardon, S. F. and J. P. Robinson (2010). Regression discontinuity designs with multiple rating-score variables. *Working Paper*.

Rossi-Hansberg, E., P.-D. Sarte, and R. Owens (2010). Housing externalities. *Journal of Political Economy 118*(3), 485–535.

Schelling, T. C. (1969). Models of segregation. *American Economic Review, Papers and Proceedings 59*(2), 488–493.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology 1*(2), 143–186.

Schelling, T. C. (1978). *Micromotives and Macrobehavior.* New York: Norton and Co.

Sherwell, A. (1897). *Life in West London.* London: Methuen and Co.

Shiode, S. (2012). Revisiting John Snow's map: network-based spatial demarcation of cholera area. *International Journal of Geographical Information Science 26*(1), 133–150.

Snow, J. (1855). *On the Mode of Communication of Cholera.* New Burlington Street, London, England: John Churchill.

Takahashi, S. (2005). Infinite horizon common interest games with perfect information. *Games and Economic Behavior 53*(2), 231–247.

The Cholera Inquiry Committee (1855). Report on the Cholera Outbreak in the Parish of St. James, Westminster, during the Autumn of 1854. Technical report, Presented to the Vestry.

Weiner, D. (1994). *Architecture and Social Reform in Late-Victorian London.* Manchester: Manchester University Press.

Whitehead, R. (1854). The Cholera in Berwick Street. Technical report, The Senior Curate of St. Luke's.

Wong, V., P. M. Steiner, and T. D. Cook (2012). Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four estimation methods. *Journal of Educational and Behavioral Statistics 38*(2), 107–141.

Zoopla (2015). Zoopla current value estimates. `http://www.zoopla.co.uk/`. [Online; accessed 20-February-2015].

Appendix A: Proofs

Label apartments in the block $i = 1, ..., n$. Define a history at a negotiation opportunity at period $t \in \mathbb{Z}_+$, denoted by h_t, as a list comprising of the time, the apartment label and the negotiation outcome (previous tenant retained, new poor tenant hired, new rich tenant hired) for realized renegotiation opportunities preceding period t, plus the apartment label for the renegotiation opportunity at t. Let H_t be the set of all time t histories as above, and let $H = \underset{t \in \mathbb{Z}_+}{\cup} H_t$. Strategies of the landlord are defined as mappings from H to {poor,rich} (where it is implicitly assumed that action poor means retaining the previous tenant if her type is poor and hiring a new poor tenant otherwise; similarly action rich means retaining the old tenant if her type is rich and hiring a new rich tenant otherwise). For every $h \in H$, let $x(h)$ be the current number of poor tenants in other apartments at the time of the negotiation associated with h.

We assume that agreed upon rents are determinded by the landlord's strategy, through the maximum rent the chosen tenant is willing to pay, given the landlord's strategy. Thus, we assume that tenants correctly foresee the landlord's actions in the future, and that they have correct expectations on how the composition of the block changes over time. The landlord chooses a strategy maximizing his expected discounted rent revenue.[55]

An alternative, and simpler way of thinking about the landlord's strategies is the following. Let $\mathcal{T}$ be the set of all possible sequences of negotiation opportunities over time, with each member of the sequence indicating the time and apartment label of the negotiation. A typical $\sqcup \in \mathcal{T}$ is of the form $(t_0, i_0), (t_1, i_1), ...$ where $t_0 = 0$ and i_0 is the label of the initially vacant apartment. We refer to (t_k, i_k) as the kth negotiation in the sequence. Then we can define the landlord's strategy as a mapping that for every negotiation of every possible sequence in $\mathcal{T}$ allocates an action from {poor,rich}, in a way that if $\sqcup, \sqcup' \in \mathcal{T}$ are such that $(t_l, i_l) = (t'_l, i'_l)$ for $l = 1, ..., k$ then the action allocated to the kth negotiation has to be the same for the two sequences (actions can only be conditioned on past events, not on future ones). Defining strategies this way has the convenient feature that the set of strategies the same for different initial compositions of tenants. In particular, given two different histories h and h', and a continuation strategy s in the game starting at h, we

---

[55] Below we show that the landlord never has an incentive to deviate from his ex ante optimal strategy, hence we do not need to assume that he can commit at $t = 0$ to follow it.

can define a sequence-equivalent strategy s' in the game starting at h' as a strategy allocating the same action as s to every negotiation of every possible negotiation sequence.

Lemma 1: Let $h \in H$ and relabel apartments in the game starting at h' such that every apartment having a poor tenant at h also has a poor tenant at h'. Let s be any strategy in the game starting at h and let s' be a sequence equivalent strategy to s in the game starting with h'. Then the payoff that s yields to the landlord given h is weakly lower than the payoff s' yields given h'.

Proof: Since $x(h) \geq x(h')$ and s' is a sequence equivalent strategy to s, for any sequence of negotiations $\sqcup$ the number of poor tenants under s is weakly higher than under s'. Hence, at any future negotiation newly hired tenants expect in any future period weakly higher number of poor neighbors under s and are ready to pay weakly lower rent. As a result, the payoff that s yields to the landlord given h is weakly lower than the payoff s' yields given h'.

Theorem 1: The landlord always has an optimal strategy of the following form: there is $x^* \in \{0, ..., n-1\}$ such that at every history $h \in H$, if $x(h) \leq x^*$ then choose rich, and if $x(h) > x^*$ then choose poor.

Proof: To simplify notation below, denote the initial history, at $t = 0$, simply as h in this proof. First note that if $x(h) = 0$ then choosing rich at h and in all future negotiations is an optimal continuation strategy, as it results in the maximum possible negotiated wage (W^r) at every negotiation of the continuation game. Moreover, if $h' \in H$ is on the path of play given the landlord's continuation strategy at h, and $x(h') = 0$ then an optimal strategy has to choose rich at h' and at all successor histories on the path of play. This is because only those strategies can maximize the landlord's expected payoff given h', and at the same time maximize the rent for rich tenants retained/hired preceding h'.

Let x^* be largest number of initial poor tenants such that whenever $x(h) \leq x^*$, there exists an optimal strategy s given h such that rich is chosen at h. As shown above, the requirement holds for $x = 0$.

Assume $x^* \geq 1$ and consider $x(h) = 1$. Assume that the landlord is playing an optimal strategy which specifies acquiring a rich tenant at h. Note that for every immediate successor history h' of h, either $x(h') = 1$ or $x(h') = 0$. As shown above, in the latter case an optimal strategy of the landlord has to choose rich at h'. Next, for all h' such that $x(h') = 1$, change the continuation strategy that s specifies at h' to s itself (with the label of the negotiated apartment at h exchanged with the label of the negotiated apartment at h'). Since s is optimal at h, and the game starting at h' is equivalent (up to relabeling apartments) to the game starting at h, the new strategy s' is optimal conditional on h' and yields weakly higher continuation payoffs at every immediate successor h' of h. For now, fix the rich rent at h at the level it would be when s is played. Then s' with the old rent at h yields a weakly higher payoff for the landlord than s. Next, we can replace continuation strategies at all h'' that are immediate successors of h' that are immediate successors of h, with $x(h'') = 1$ to s. Analogous arguments as before establish that s'' is optimal conditional on h'' and yields weakly higher continuation payoffs at every immediate successor h'' of h' than s'. For now, keep rich rent levels agreed upon prior to h'' unchanged. Then s'' with the old rent levels prior to h'' yields a weakly higher payoff for the landlord than s'. Iterating the argument establishes that a continuation strategy that for any successor h' of h with $x(h) = 1$ chooses rich, fixing previous rich rents, yields a weakly higher payoff than s. Now revisit all the rents that were fixed at different steps of the iteration. Conditional on any history, the rich rent is maximized if landlord plays always rich strategy from that point on. Therefore all the rents fixed before can only increase. Hence, a continuation strategy that for any successor h' of h with $x(h) = 1$ chooses rich yields a weakly higher payoff than s, therefore it is optimal. Moreover, for any $h \in H$, there is an optimal strategy that for any h' that is a successor of h and satisfies $x(h') \in \{0, 1\}$, it specifies choosing rich at h', since the latter is the optimal continuation strategy at h' and among all continuation strategies at h', it maximizes the rent for rich tenants retained/hired preceding h'.

Iterating the previous argument establishes that there is an optimal strategy of the landlord, that for any h' that is a successor of h and statisfies $x(h') \in \{0, ..., \}$, specifies choosing rich at h'.

Assume next that in every optimal strategy s given h, poor is chosen at h (this in particular requires $x(h) > x^*$), but the always poor strategy is not optimal given h. Then there exists a successor $h' \in H$ such that for every history h'' preceding h' poor is chosen, but at h' rich is chosen. Note that s has to specify a continuation strategy at h' that is optimal given h', since at every history preceding h' a poor type is hired/retained, hence the rent obtained by the landlord is independent of the continuation strategy at h'. But below we show that it cannot be that s is optimal given both h and h', leading to a contradiction.

Let $W(x)$ be the expected discounted present value of all rents from rental agreements negotiated at or after time 0 when the initial number of poor tenants is x and the landlord chooses an optimal strategy.

$W(x)$ is the sum of the rent that is received from the tenant currently being hired plus the continuation utility received from future negotiated rents, given an optimal strategy. Assume that there is an optimal strategy for the owner to first hire a poor person, but $W(x)$ is greater than what he could get from an always poor strategy, which is equivalent to $W(x) > g^* = \frac{(1-\delta(1-q))W^p}{(1-\delta)(1-\delta(1-\frac{q}{n}))}$. Then the continuation utility after current hire:

$$\begin{aligned} W(x) - \frac{W^p}{1-\delta(1-\frac{q}{n})} \leq \delta(1-q)\left(W(x) - \frac{W^p}{1-\delta(1-\frac{q}{n})}\right) + \\ + \delta q \frac{x+1}{n} W(x) + \delta q \left(1 - \frac{x+1}{n}\right) W(x+1) \end{aligned}$$

From Lemma 1 we know that $W(x+1) > W(x)$ cannot be the case because if at the game starting with x poor the owner uses a sequence equivalent strategy to an optimal strategy of the game starting with $x+1$ poor, his payoffs (from noninitial rentors) are weakly higher. But if $W(x+1) \leq W(x)$, then from the inequality for the continuation utility we have $(1-\delta)W(x) \leq \frac{(1-\delta(1-q))W^p}{1-\delta(1-\frac{q}{n})}$ or, equivalently, $W(x) \leq g^*$, which contraddicts our assumption. This leads to a contradiction, establishing that $W(x) = g^*$ and if it is optimal to start with hiring a poor, then always poor must be an optimal strategy.

The above argument establishes that if in every optimal strategy s given h, poor is chosen at h then the always poor strategy is optimal given h. In particular, the always poor strategy is optimal if $x(h) = x^* + 1$ (provided $x^* < n - 1$). Now assume that $x^* < n - 2$, $x(h) = x^* + 2$, and there exists an optimal strategy s given h such that rich is chosen at h. But Lemma 1 establishes that for a history h' with $x(h') = x^* + 1$, the game starting at h' has a strategy that chooses rich at h', and yields a weakly higher expected payoff to the landlord than s does in the game starting at h. Moreover, note that the always poor strategy yields the same expected payoff to the landlord in both games. But then there exists an optimal strategy in the game starting at h' that chooses rich at h', contradicting the definition of x^*. Hence $x(h) = x^* + 2$ implies that there is an optimal strategy given h such that poor is chosen at every h' with $x(h') > x(h)$. Iterating the above argument establishes the same conclusion for any h such that $x(h) > x^*$.

Putting together the above-derived results yields that the strategy that specifies choosing rich at a history h' iff $x(h') \leq x^*$ is optimal given h, for any $x(h)$. ■

Note that the above optimal strategy of the landlord is optimal not only given h, but also given any successor history h'. Therefore the landlord does not need to be able to commit to follow the strategy - it is in his own interest to stick to it. Also note that the strategy implies either always retaining/hiring poor types or always retaining/hiring rich types, since if at the initial history a rich type is hired then the number of poor tenants is wekly lower at all subsequent negotiations, while if at the initial history a poor type is hired then the number of poor tenants is wekly lower at all subsequent negotiations.

Proof of Proposition 1

Consider a rich tenant, who pays r per period. If he realizes his outside option, he gets $V(out) = -\frac{W^r}{1-\delta}$. Let V_k denotes the expected continuation utility of a rich tenant renting an apartment for a general fixed r, given k current poor tenants, assuming that the landlord is following the always rich strategy. If the tenant has no poor neighbours, then next period three situations are possible: with probability $1-q$ no changes; with probability $q\frac{n-1}{n}$ one rich neighbour's contract expires; with probability $\frac{q}{n}$ the tenant's contract expires, in which case his continuation utility is equal to $V(out)$. Hence, we can write:

$$V_0 = -(r+c_0^r) + \delta[(1-q)V_0 + q\frac{n-1}{n}V_0 + \frac{q}{n}V(out)]$$

$$V_0 = \frac{-(r+c_0^r) + \delta\frac{q}{n}V(out)}{1-\delta(1-\frac{q}{n})} = \frac{-(r+c_0^r) - \frac{q}{n}\frac{\delta}{1-\delta}W^r}{1-\delta(1-\frac{q}{n})} \tag{7}$$

If the tenant has $k \geq 1$ poor neighbours, then next period four situations are possible: with probability $1-q\frac{k+1}{n}$ no changes; with probability $q\frac{k}{n}$ one poor neighbour is replaced by a rich one; with probability $\frac{q}{n}x$ the tenant's contract expires and her continuation utility is equal to $V(out)$. Hence, we get:

$$V_k = -(r+c_k^r) + \delta[(1-q\frac{k+1}{n})V_k + q\frac{k}{n}V_{k-1} + \frac{q}{n}V(out)] \tag{8}$$

$$V_k = \frac{\delta q\frac{k}{n}}{1-\delta(1-q\frac{k+1}{n})}V_{k-1} + \frac{\delta\frac{q}{n}V(out) - (r+c_k^r)}{1-\delta(1-q\frac{k+1}{n})}$$

Iterating, we obtain:

$$V_k = \frac{\delta \frac{q}{n} V(out) - r}{1 - \delta(1 - \frac{q}{n})} - \sum_{i=0}^{k} \frac{\frac{k!}{i!}(\delta \frac{q}{n})^{k-i}}{\prod\limits_{j=i}^{k}(1 - \delta(1 - q\frac{j+1}{n}))} c_i^r$$

The apartment owner chooses rent r_x by making a rich tenant indifferent between renting and outside option: $V_x = V(out)$. Hence:

$$r_x = W_r - \Big(1 - \delta(1 - \frac{q}{n})\Big) \sum_{i=0}^{x} \frac{\frac{x!}{i!}(\delta \frac{q}{n})^{x-i}}{\prod\limits_{j=i}^{x}(1 - \delta(1 - q\frac{j+1}{n}))} c_i = W_r - \sum_{i=0}^{x} a_{ix} c_i^r \qquad (9)$$

$$a_{ix} = \Big(1 - \delta(1 - \frac{q}{n})\Big) \frac{\frac{x!}{i!}(\delta \frac{q}{n})^{x-i}}{\prod\limits_{j=i}^{x}(1 - \delta(1 - q\frac{j+1}{n}))} \qquad (10)$$

Consider the apartment owner, who has x poor tenants and follows the always rich strategy. His expected utility $U_{rich}(S_r, x)$ can be divided into the expected payoff from contacts agreed upon before time 0, $U_{curr}(S_r)$, and the expected payoff from contracts negotiated time 0 on, under the always rich strategy, f_x. The latter consists of the expected payoff from the time 0 contract and the expected payoff from all future contracts, denoted by h_x.

$$U_{rich}(S_r, x) = U_{curr}(S_r) + f_x = \frac{S_r}{1 - \delta(1 - \frac{q}{n})} + \frac{r_x}{1 - \delta(1 - \frac{q}{n})} + h_x$$

$$f_0 = \frac{W^r - c_0^r}{1 - \delta(1 - \frac{q}{n})} + \sum_{i=1}^{\infty} \delta^i q \frac{W^r - c_0^r}{1 - \delta(1 - \frac{q}{n})} = \frac{(1 - \delta + \delta q)(W^r - c_0^r)}{(1 - \delta)(1 - \delta(1 - \frac{q}{n}))}$$

As there are $k \geq 1$ poor tenants in the current period, then next period with probability $q(1 - \frac{k}{n})$ a rich tenant's rent gets renegotiated to r_k, and with probability $q\frac{k}{n}$ a rich tenant replaces a poor one with a negotiated rent r_{k-1}. Therefore:

$$h_k = \delta[(1 - q)h_k + q(1 - \frac{k}{n})f_k + q\frac{k}{n}f_{k-1}]$$

$$(1 - \delta(1 - q))\left(f_k - \frac{r_k}{1 - \delta(1 - \frac{q}{n})}\right) = \delta q(1 - \frac{k}{n})f_k + \delta q\frac{k}{n}f_{k-1}$$

$$f_k = \frac{\delta q \frac{k}{n}}{1 - \delta(1 - q\frac{k}{n})} f_{k-1} + \frac{1 - \delta(1 - q)}{(1 - \delta(1 - \frac{q}{n}))(1 - \delta(1 - q\frac{k}{n}))} r_k$$

Solving the difference equation, we get:

$$f_k = \frac{1-\delta+\delta q}{(1-\delta)(1-\delta(1-\frac{q}{n}))} W^r - \sum_{i=0}^{k} b_{ik} c_i \tag{11}$$

$$b_{ik} = (1-\delta(1-q)) \frac{\frac{k!}{i!}(k+1-i)(\delta\frac{q}{n})^{k-i}}{\prod\limits_{m=i}^{k+1}(1-\delta(1-q\frac{m}{n}))} \tag{12}$$

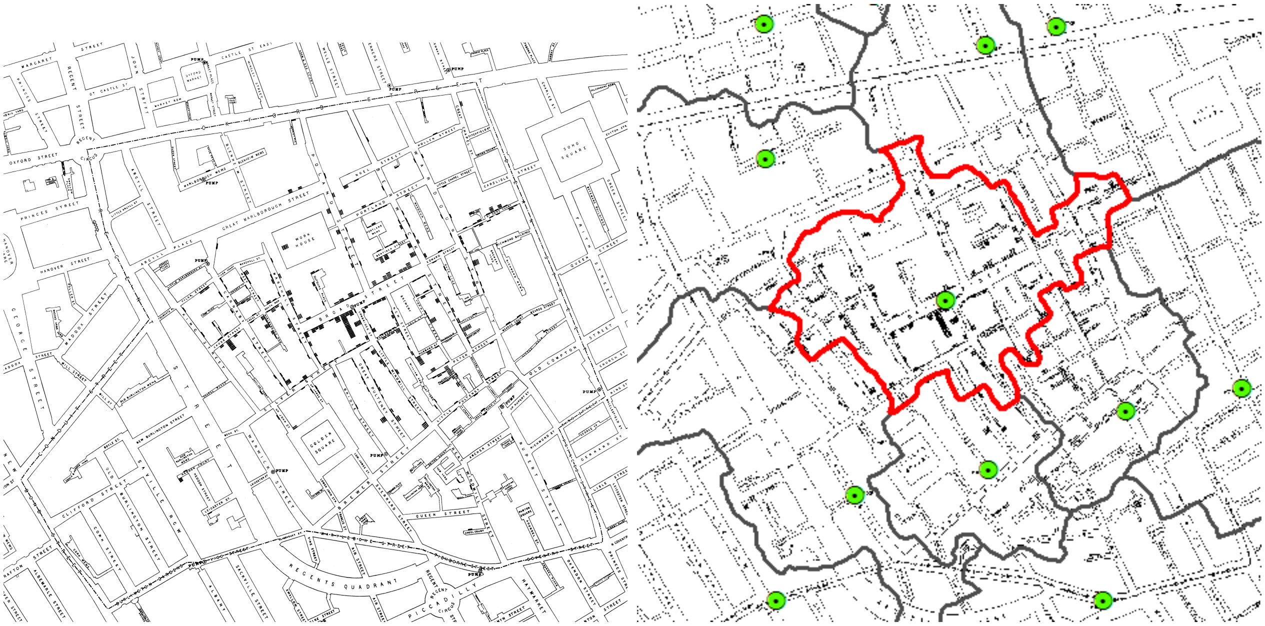
Now consider the always poor strategy. The owner's expected utility U_{poor} can be divided into expected payments from current contracts negotiated before time 0, $U_{curr}(S_r)$, and the expected payments from contracts negotiated at time 0 on when the landlord is playing the always poor strategy, denoted by g.

$$\begin{aligned} U_{poor}(S_r, x) &= U_{curr}(S_r) + g \\ g &= \frac{(1-\delta+\delta q)}{(1-\delta)(1-\delta(1-\frac{q}{n}))} W^p \end{aligned} \tag{13}$$

We can conclude that the apartment owner, having x poor tenants, prefers the always rich strategy to the always poor strategy if $f_x > g$ or, equivalently,

$$W^r - W^p > (1-\delta)\left(1-\delta(1-\frac{q}{n})\right)\left[\sum_{i=0}^{x} \frac{\frac{x!}{i!}(x+1-i)(\delta\frac{q}{n})^{x-i}}{\prod\limits_{m=i}^{x+1}(1-\delta(1-q\frac{m}{n}))} c_i^r\right]$$

Figures and Tables

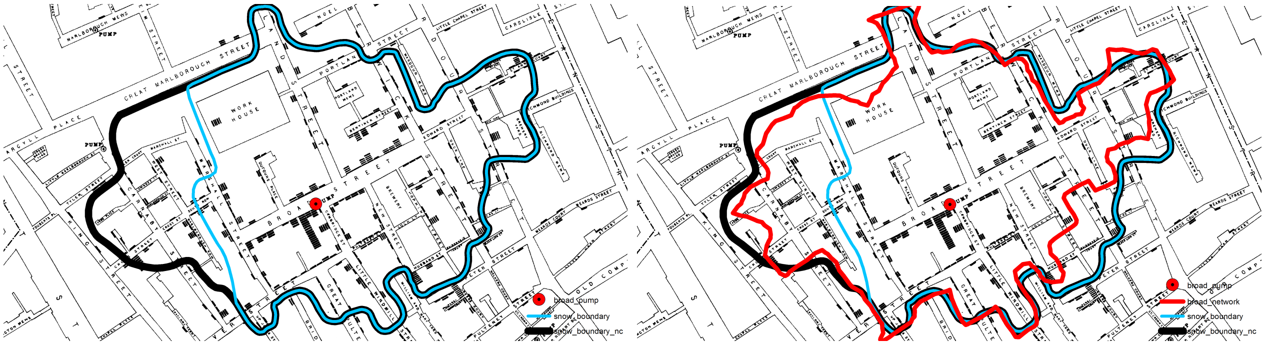


(a) John Snow's 1854 cholera map

(b) Catchment areas

Figure 1: John Snow's 1854 Cholera Map with Pump's Catchment Areas

Notes: Green dots indicate the location of a pump. Broad Street pump catchment area highlighted in red. Each catchment area is defined by a network Voronoi polygon.



(a) John Snow's boundaries

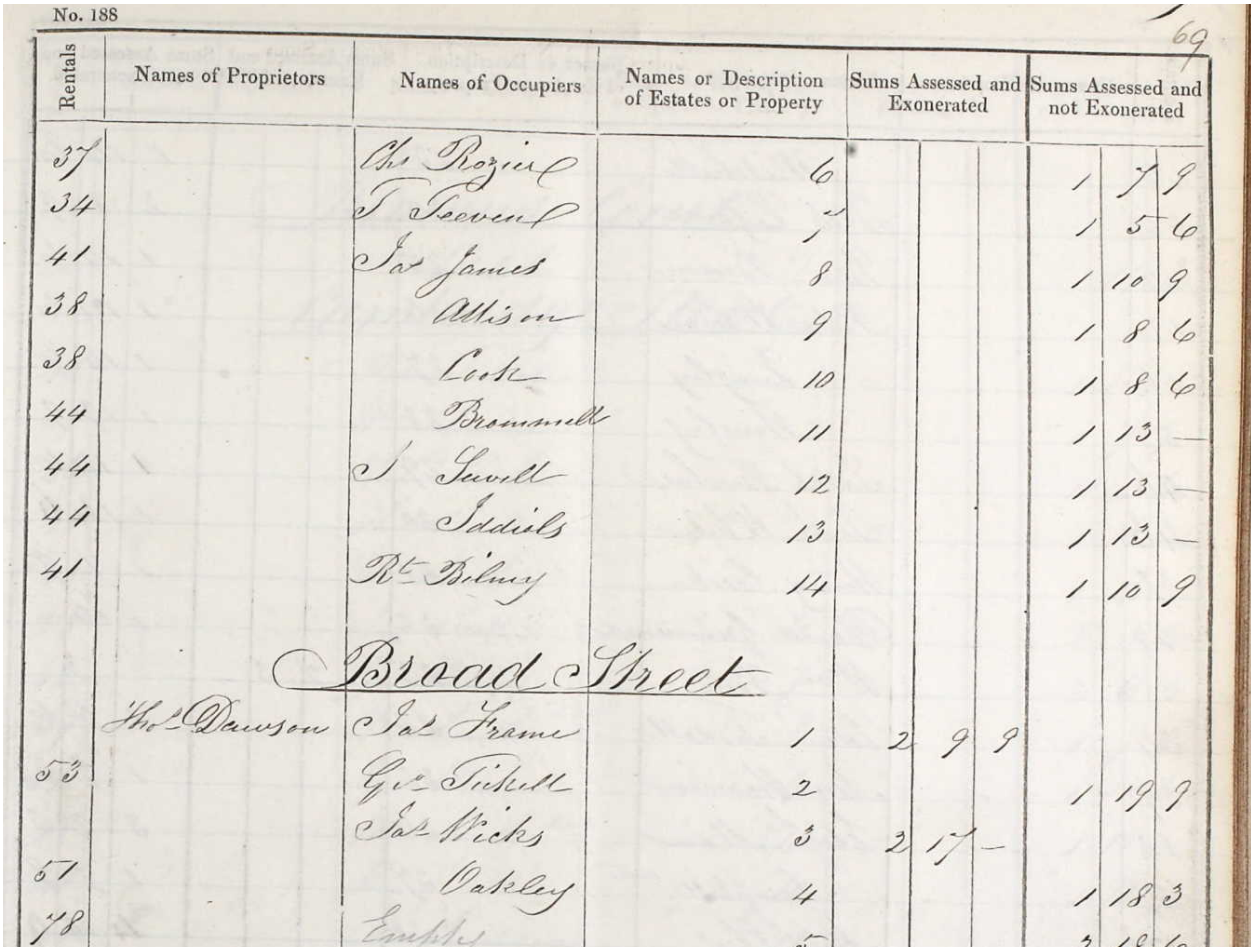
(b) John Snow's boundaries and Voronoi boundary

Figure 2: John Snow's BSP Boundaries

Notes: Boundary colored blue depicts John Snow's original boundary. Boundary in black is a modification of John Snow's original boundary that excludes the pump on Little Marlborough St. Boundary colored red depicts the shortest walking distance boundary used in previous specifications (i.e., Voronoi boundary)

Figure 3: Cholera Inquiry Committee (1855) Cholera Deaths Map

Notes: Black bars represent a cholera death

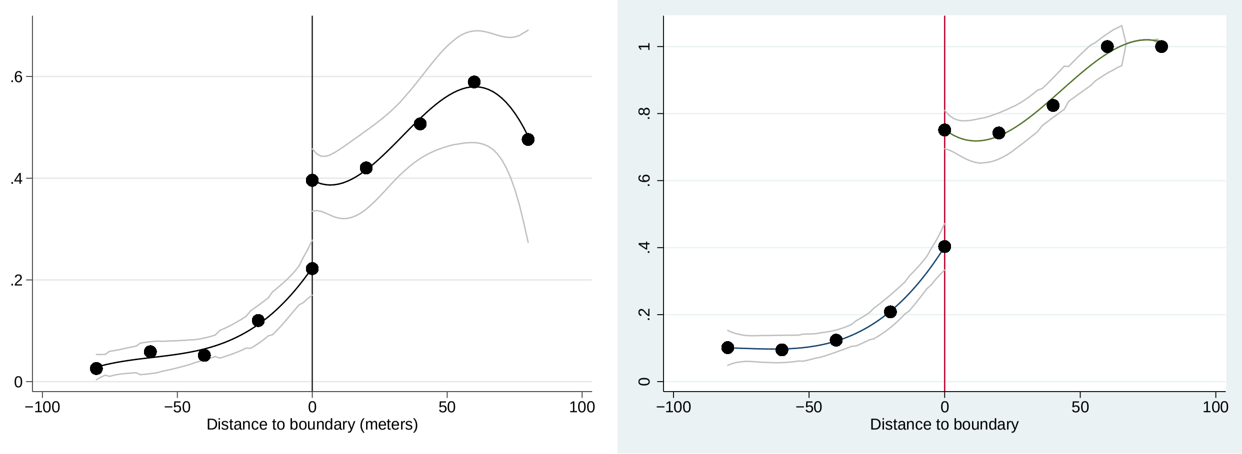


No. 188

69

Rentals	Names of Proprietors	Names of Occupiers	Names or Description of Estates or Property	Sums Assessed and Exonerated	Sums Assessed and not Exonerated
37		Chs Rozier	6		1 7 9
34		T Peever	7		1 5 6
41		Jas James	8		1 10 9
38		Allison	9		1 8 6
38		Cook	10		1 8 6
44		Brommell	11		1 13 —
44		J Sewell	12		1 13 —
44		Iddiols	13		1 13 —
41		Rt Bilney	14		1 10 9
		Broad Street			
	Thos Dawson	Jas Frame	1	2 9 9	
53		Geo Tickell	2		1 19 9
		Jas Wicks	3	2 17 —	
57		Oakley	4		1 18 3
78		Empty	5		

Figure 4: Land Tax records, Broad Street, 1853

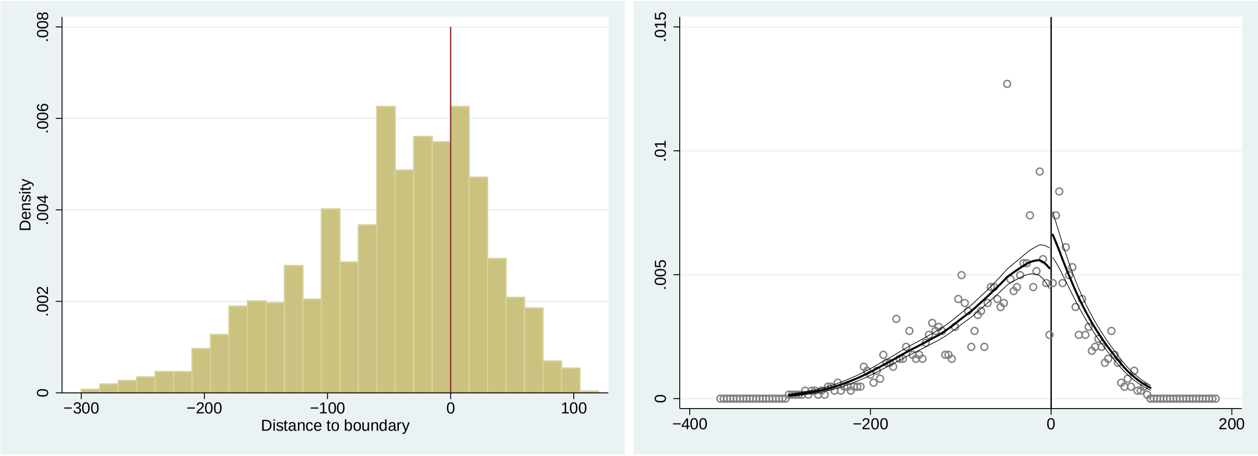


(a) House has at least one death

(b) At least 25 % of houses in block have a death

Figure 5: Cholera Deaths and BSP Boundary (1854)

Notes: Each point represents the average value of the specified variable for distance-to-boundary bins that are 20 meters wide. Negative/positive values of distance give the distance of houses inside/outside BSP catchment area, respectively. Solid line trends are the predicted values and corresponding 95 percent confidence intervals from a regression of the specified variable on a third degree polynomial in distance to the boundary.

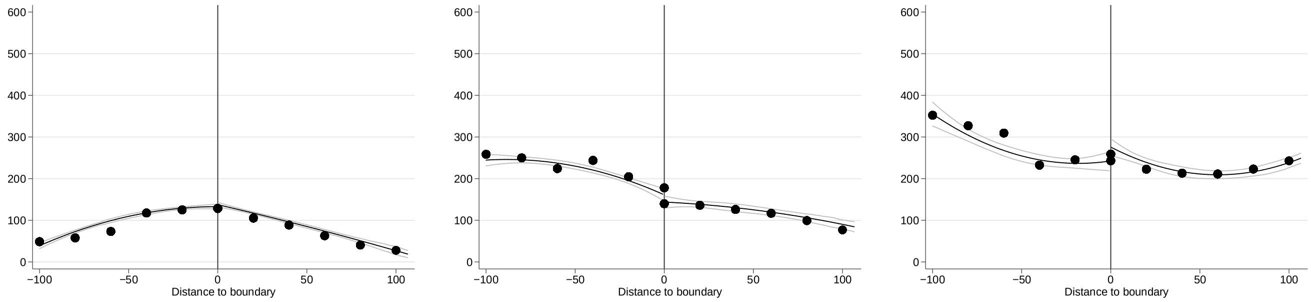


(a) Distance to BSP boundary histogram

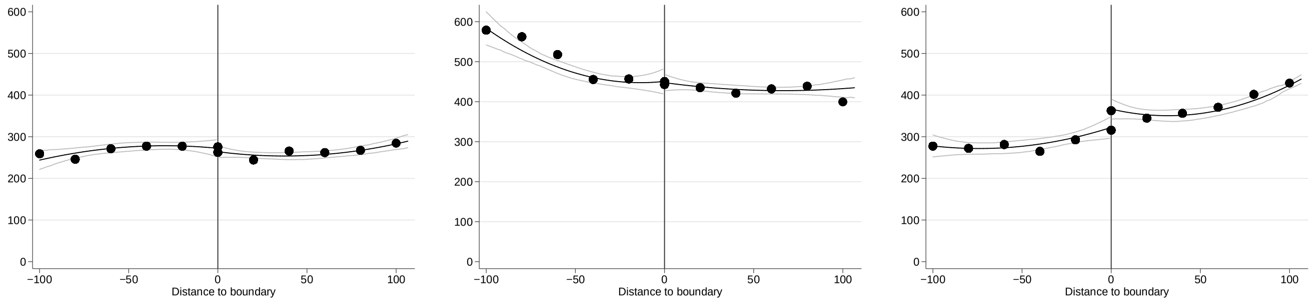
(b) McCrary's (2008) test

Figure 6: Histogram and Density of Forcing Variable (Distance to BSP boundary)

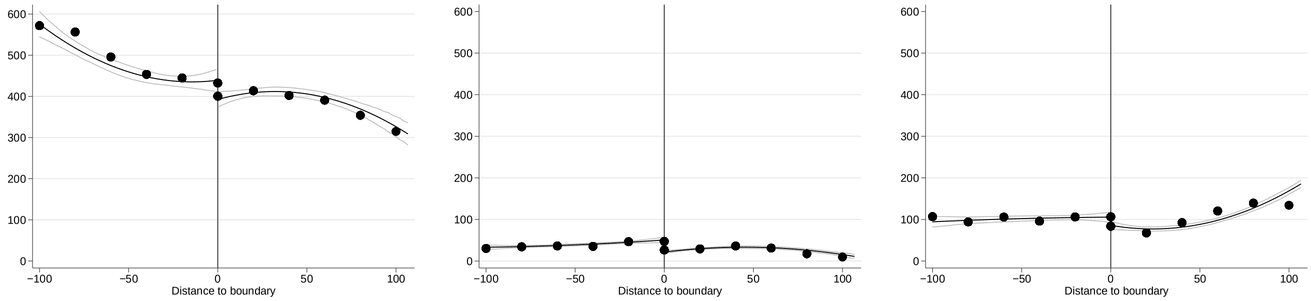
Notes: "Distance to boundary" refers to the distance between a house and the closest point in the BSP boundary. Positive/negative values of distance give the distance of houses inside/outside BSP area respectively. Distance is measured in meters. Bins width is 15 meters. Solid vertical line represents the treatment boundary.



(a) Closest pump (b) Soho centroid (c) Pressumed plage pit

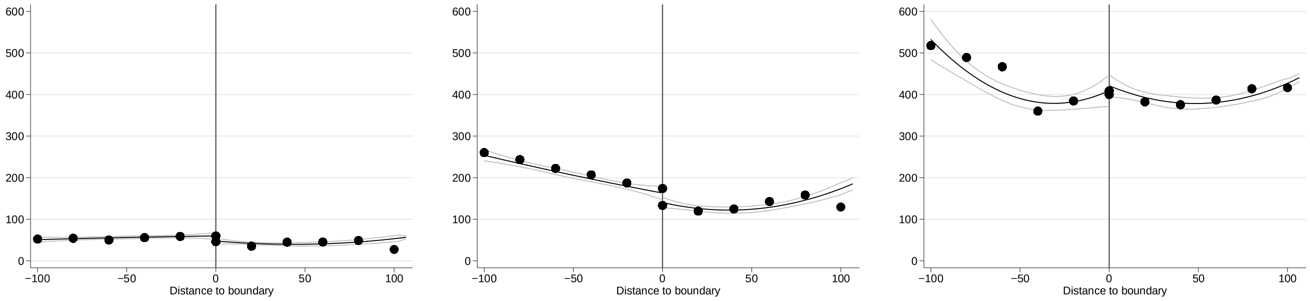


(d) Public square (e) Police station (f) Fire station

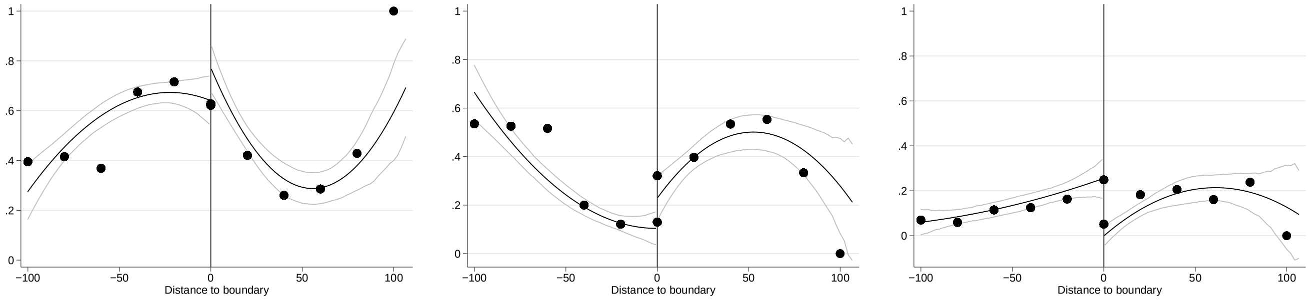


(g) Theater (h) Pub (i) Urinal

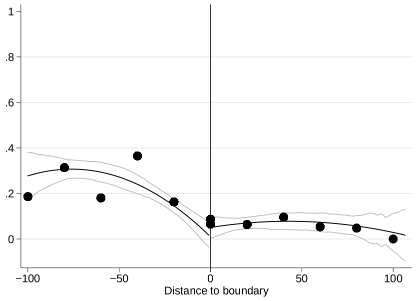
Figure 7: Covariate RD Plots (1853) - *Continues*



(a) Sewer vent (b) Primary school (c) Bank



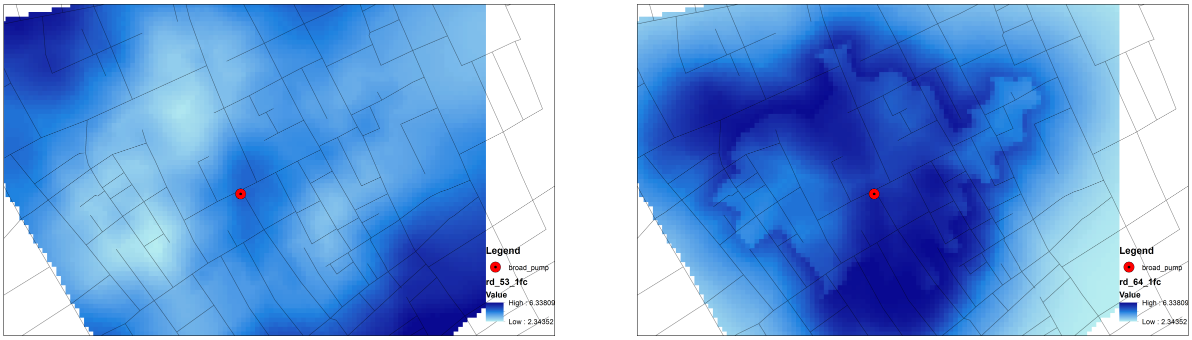
(d) Old/Existing sewer (e) New sewer (f) No sewer access



(g) Tax exonerated

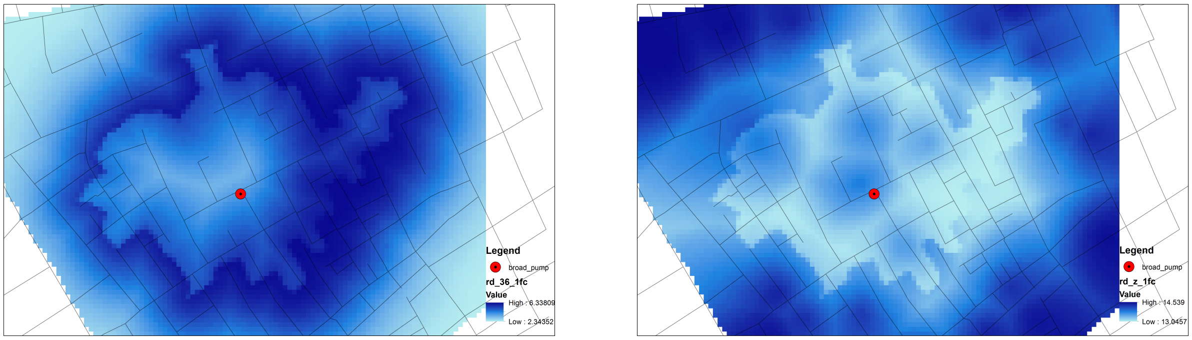
Covariate RD Plots (1853) - *Continued*

Notes: Solid dots give the average value of the specified variable for houses falling within 20 meter distance bins. Dots are plotted at the start of the bin (i.e. the dot representing the average for houses in the 0-20 meter bin is located at 0.). "Distance to boundary" refers to the distance between a house and the closest point in the BSP boundary. Distance is measured in meters. The solid vertical line represents the BSP boundary. Negative/positive values of distance give the distance of houses inside/outside BSP area respectively. The solid line trends are the predicted values from a regression of the specified variable on a second degree polynomial in distance to the boundary that uses a rectangular kernel and a bandwidth of 200 meters.



(a) Rental price (1853)

(b) Rental price (1864)

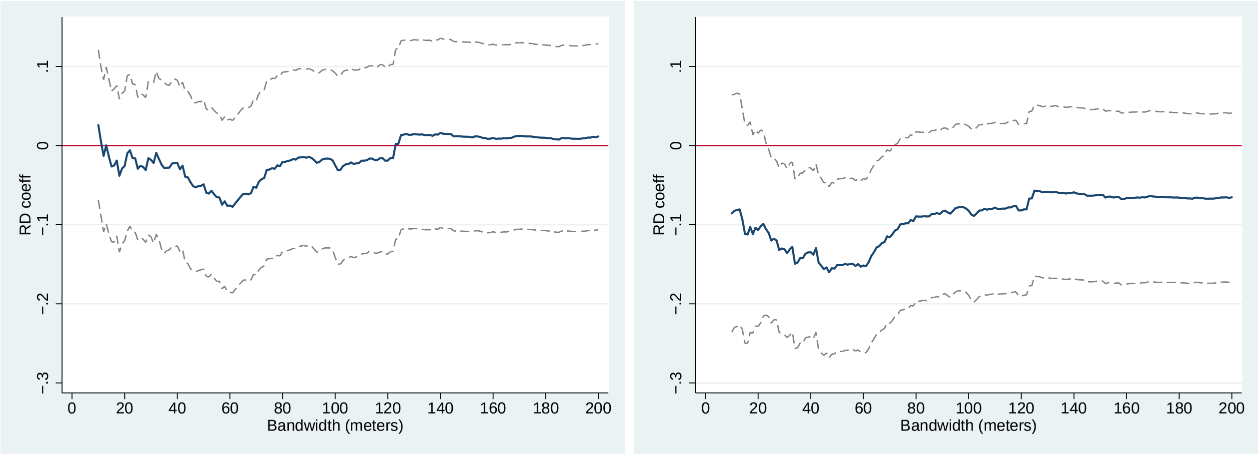


(c) Rental price (1936)

(d) House price (1995-2013)

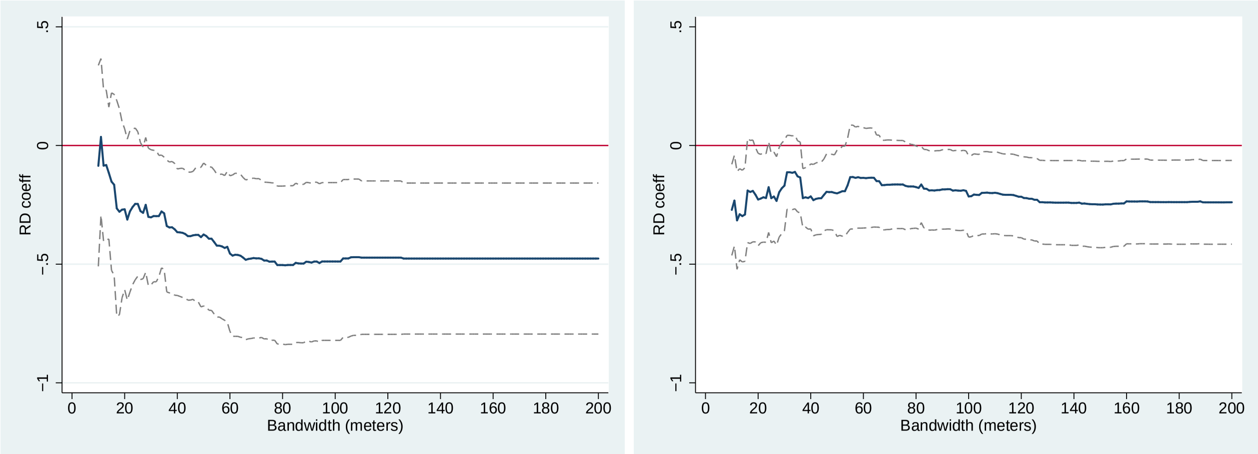
Figure 8: RD Plots, Outcome Variables

Notes: Monochromatic scale gives the predicted values from an RD model using a first degree polynomial in distance to BSP boundary and baseline controls. Color scale is smoothed using the Raster Stretch tool in ArcGIS. Panels (a), (b), and (c) share the same color scale. Panel (d) uses a different scale since model uses house price instead of rental price as outcome variable.



(a) RD coefficient (1853)

(b) RD coefficient (1864)

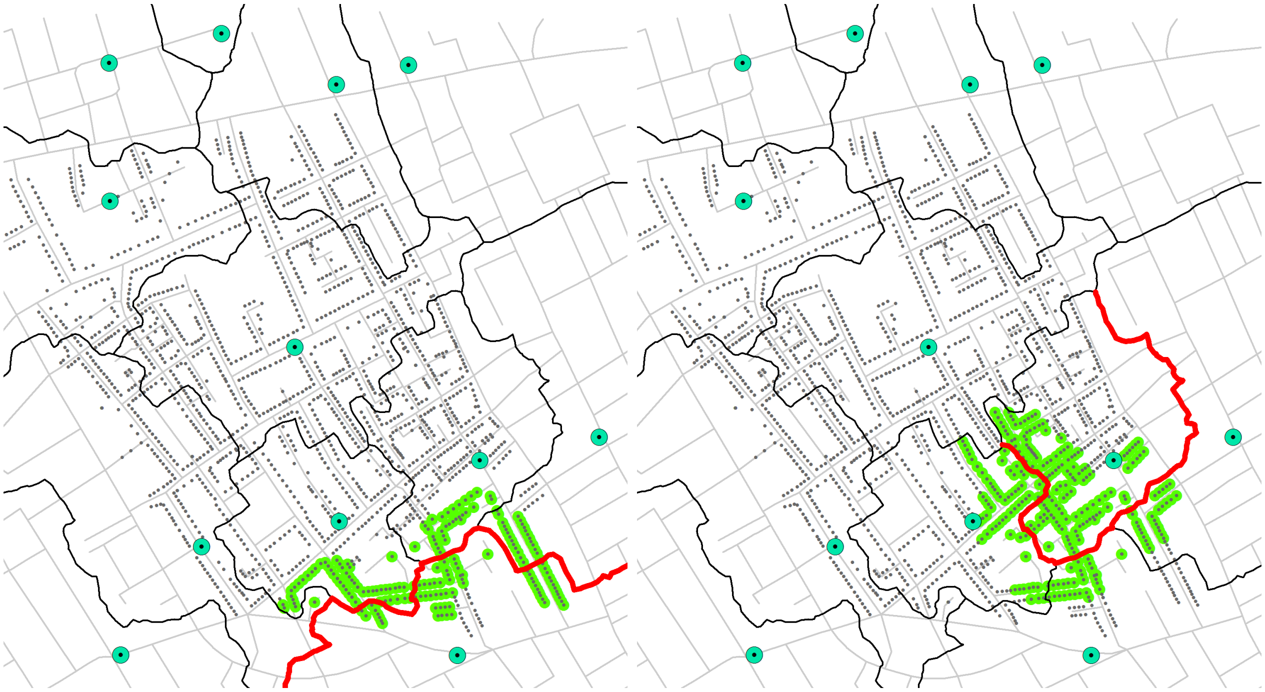


(c) RD coefficient (1936)

(d) RD coefficient (1995-2013)

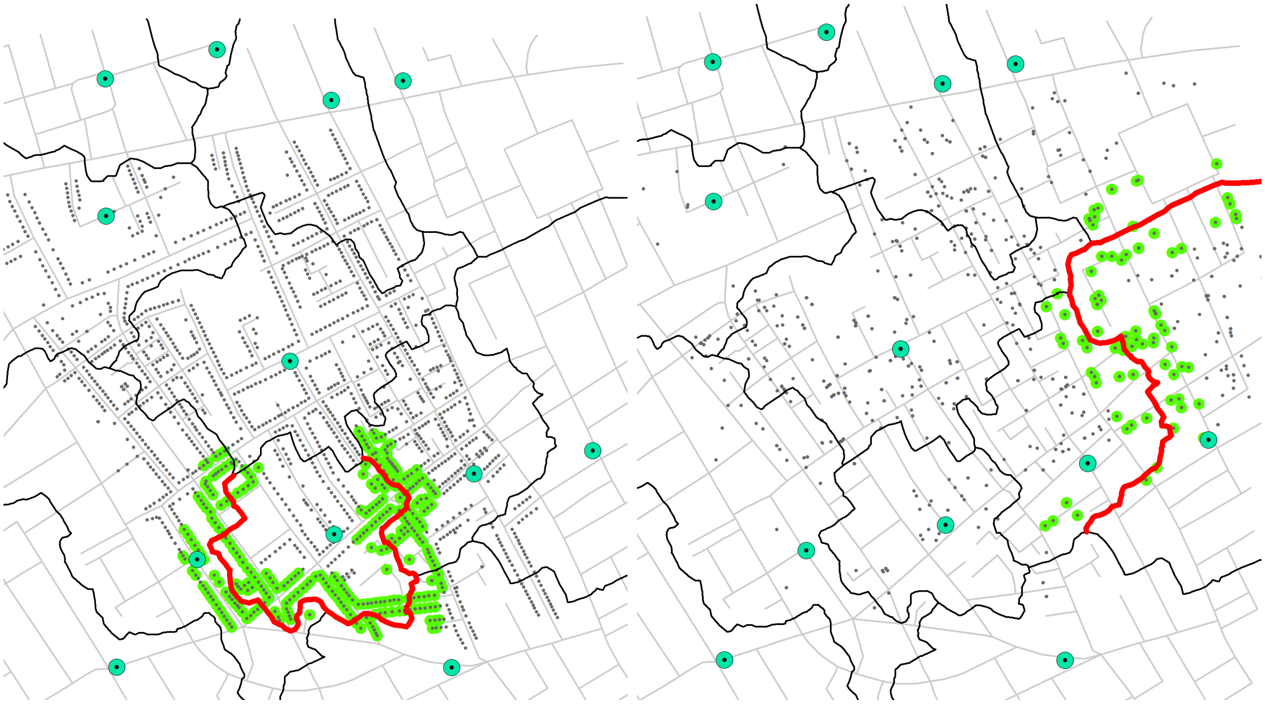
Figure 9: Bandwidth Sensitivity

Notes: Dashed lines represent 90 percent confidence intervals. “RD coefficient” refers to the coefficient estimate of BSP in the main estimation equation.



(a) Boundary 1

(b) Boundary 2

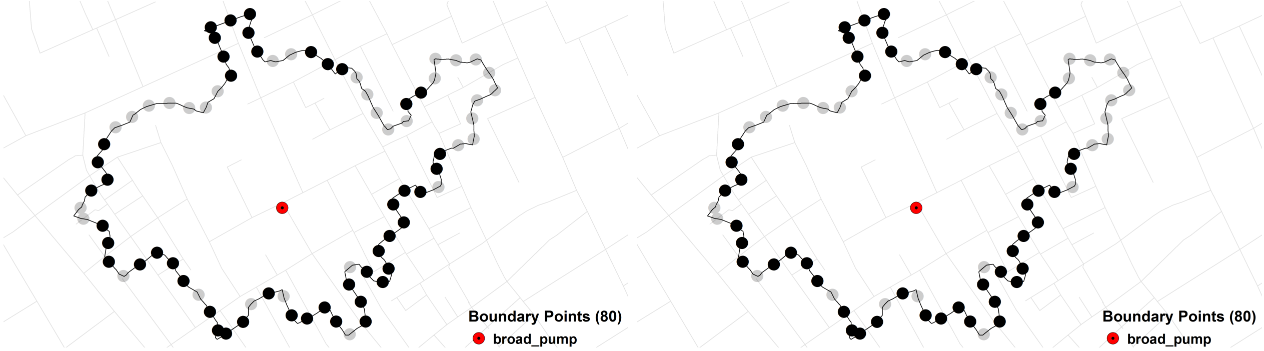


(c) Boundary 3

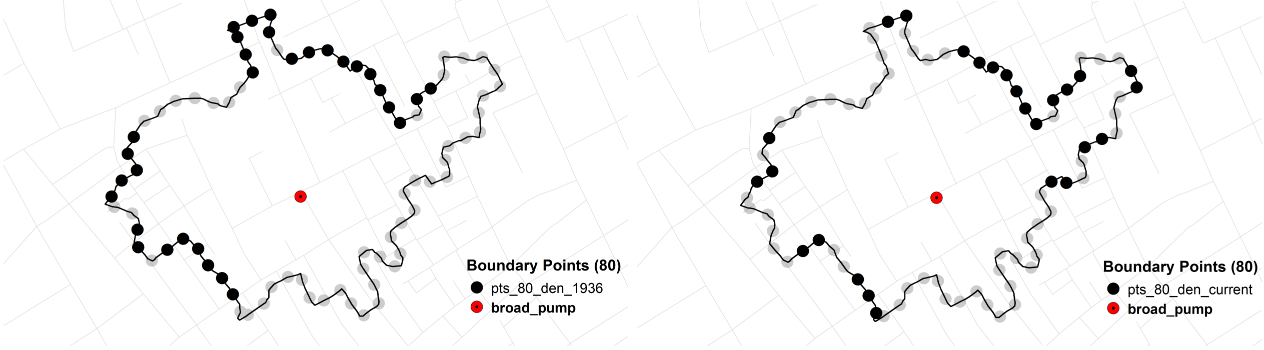
(d) Boundary 4

Figure 10: False Treatment Boundaries and Estimation Samples

Notes: False boundaries are selected based on sample availability. Observations inside BSP were excluded from the analysis. Highlighted observations fall inside the optimal bandwidth used for the corresponding RD analysis. Optimal bandwidth is determined as in Imbens and Kalyanaraman (2012). The resulting bandwidths are: 37m for Boundary 1, 50m for Boundary 2 in 1853 and 1864 sample, 55m for Boundary 2 in current sample (not shown), 34m for Boundary 3, 55m for Boundary 4.



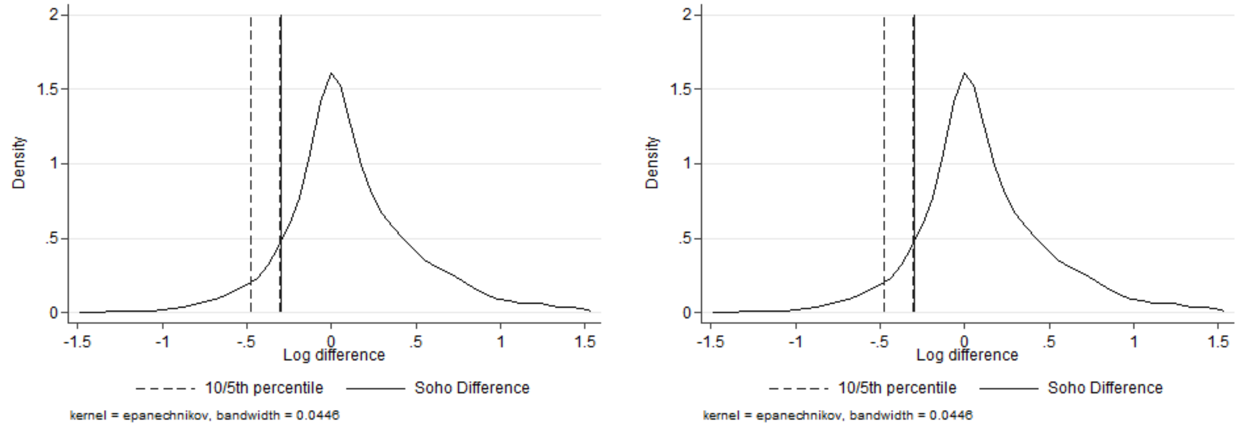
(a) Boundary points, 1853 (b) Boundary points, 1864



(c) Boundary points, 1936 (d) Boundary points, 1995-2013, 2015

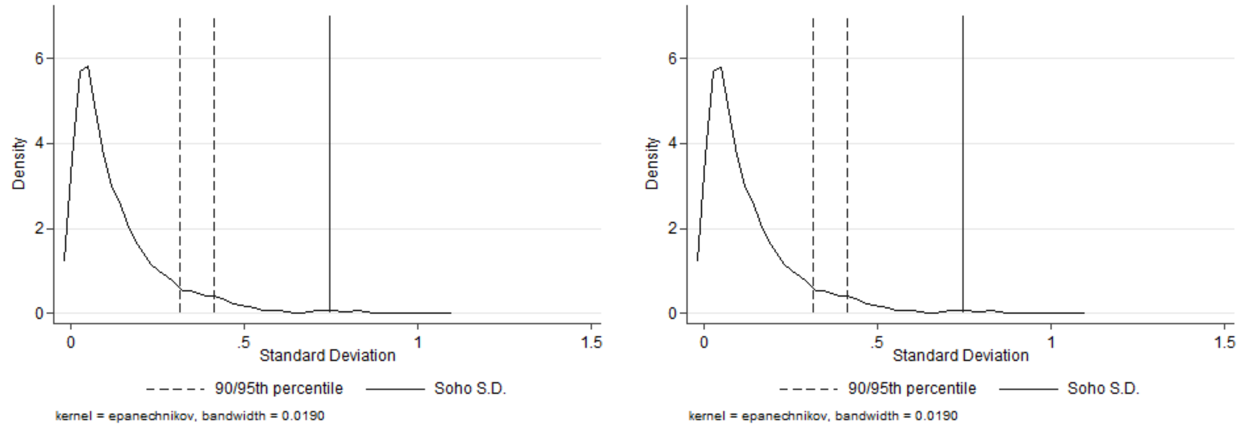
Figure 11: Boundary Points $\mathbf{b}_j$ (80)

Notes: Black dots indicate the location of boundary points for which we estimate a conditional treatment effect. Gray dots indicate the points for which we are unable to obtain an effect due to lack of observations close the BSP boundary



(a) All Prices

(b) Prices within range observed in Soho



(c) All Prices

(d) Prices within range observed in Soho

Figure 12: Difference and Standard Deviation of Log House Prices for Neighboring Blocks, London 1995-2013

Notes: Histogram of log price differences and standard deviations of log price for houses in adjacent blocks for the Greater London area. Dashed lines give the 5th and 10th percentile of the distribution in the case of the log difference (panels (a) and (b)) and the 90th and 95th percentile in the case of the distribution of standard deviations (panels (c) and (d)). Solid line gives the standard deviation of log house prices observed in the Soho area. Adjacent blocks are determined using a grid covering the Greater London area with each graticule cell having length equal to the average block size observed inside the Soho area. For a given cell, adjacent cells are defined as the four cells adjacent to each side of that cell. Sample excludes cells within the Soho area and cells intersected by a primary road, railway, monorail, river, or canal and cells with a single observation. The grid is determined using the "Create Fishnet" tool in ArcGIS.

Table 1: House Characteristics (1853)

	Full sample			Within 100 m			Opt. Bandwidth	
	Inside (1)	Outside (2)	S.E. (3)	Inside (4)	Outside (5)	S.E. (6)	RD (7)	S.E. (8)
Rental characteristics:								
Rental price (in logs)	3.723	3.780	(0.059)	3.719	3.743	(0.067)	0.126	(0.113)
Tax assessed (in logs)	0.456	0.518	(0.058)	0.453	0.496	(0.065)	0.110	(0.114)
Tax exonerated (yes = 1)	0.067	0.230	(0.040)***	0.067	0.220	(0.052)***	-0.005	(0.059)
Sewer access:								
Old/Existing	0.477	0.564	(0.085)	0.473	0.589	(0.091)	0.082	(0.135)
New sewer	0.396	0.276	(0.081)	0.399	0.260	(0.084)*	0.116	(0.105)
No access	0.127	0.159	(0.055)	0.128	0.151	(0.063)	-0.198*	(0.100)
Distance (m/100) to:								
Closest pump	1.046	0.957	(0.079)	1.052	1.066	(0.093)	0.111	(0.079)
Soho centroid	1.320	2.471	(0.119)***	1.325	2.177	(0.131)***	-0.104	(0.202)
Presumed plague pit	2.359	3.135	(0.224)***	2.358	2.630	(0.215)	0.260	(0.301)
Public square	2.586	2.717	(0.135)	2.584	2.715	(0.138)	-0.109	(0.189)
Church	1.322	1.712	(0.129)***	1.323	1.609	(0.141)**	0.071	(0.163)
Police station	4.361	5.413	(0.261)***	4.364	4.797	(0.221)*	-0.078	(0.438)
Fire station	3.603	2.664	(0.187)***	3.597	2.864	(0.227)***	0.452	(0.353)
Theater	4.004	5.306	(0.232)***	4.011	4.679	(0.219)***	-0.397	(0.335)
Pub	0.286	0.408	(0.037)***	0.287	0.406	(0.046)**	-0.184	(0.111)
Urinal	0.878	1.122	(0.087)***	0.874	1.019	(0.088)	-0.144	(0.141)
Sewer vent	0.429	0.555	(0.048)***	0.431	0.563	(0.050)***	-0.023	(0.100)
Primary school	1.306	2.474	(0.132)***	1.306	2.023	(0.109)***	-0.152	(0.195)
Bank	3.949	4.694	(0.316)**	3.947	4.095	(0.315)	-0.005	(0.520)
Observations	495	1230		491	815		534	

Notes: Columns (1), (2), (4), and (5) give the mean of the corresponding variable. Columns (3) and(6) give the clustered standard error at the block level for the difference in means. "Inside" and "Outside" indicate whether a house is inside or outside the BSP area respectively. Columns (7) and (8) give the estimated coefficient and standard error for the RD specification that uses the corresponding variable as its outcome. Optimal bandwidth is determined as in Imbens and Kalyanaraman (2012) using a triangular kernel. Column (7) uses the minimum of all optimal bandwidths (27.50 meters). *, **, and *** indicate 10, 5, and 1 percent significance respectively.

Table 2: Boundary Effects on Rental Prices, 1853

	Log Rental Price, 1853				
	Boundary RD	One-dimensional RD			
	Averaged Conditional Treatment Effects (1)	Optimal Band Parametric Form (2)	Narrow Band Parametric Form (3)	Optimal Band Cluster by Street (4)	Segment FE (5)
Inside BSP area	-0.0373	-0.0401	-0.0162	-0.0401	-0.0293
	(0.0542)	(0.0654)	(0.0560)	(0.0676)	(0.0666)
Distance to BSP boundary		-0.829	0.0669	-0.829	-0.997
		(1.320)	(0.242)	(1.467)	(1.279)
(Distance to BSP boundary)^2		2.218	-0.0531	2.218	2.961
		(8.068)	(0.238)	(8.875)	(7.961)
(Distance to BSP boundary)^3		0.807	0.0306	0.807	0.213
		(13.31)	(0.0568)	(14.55)	(13.26)
Observations	.	674	426	674	674
Boundary points	47				
Mean outside BSP area	53.31	54.4	45.71	54.4	54.4
Bandwidth (meters)	57	44.03	24	44.03	44.03

Notes: Clustered standard errors shown in parenthesis. Standard errors in Column (1) obtained via bootstrap with 1,000 replications. Bandwidth chosen optimally as in Imbens and Kalyanaraman (2012). Analysis in Column (1) uses minimum of all optimal bandwidths for each boundary point. Since sample size used in analysis at each boundary point varies, Column (1) does not report a specific number of observations. Boundary points are chosen randomly along the boundary and indicate the number of points at which the treatment effect is evaluated. Analysis restricted to boundary points with at least one observation on each side of the BSP boundary within the optimal bandwidth (Boundary Positivity Assumption). Columns (2), (3), and (5) use blocks as clusters. Column (4) uses street clusters. Narrow bandwidth is defined as the smallest bandwidth for which the RD coefficient is statistically significant. Segments in Column (5) are determined by dividing the BSP boundary into five segments of equal length. An observation is assigned to a segment based on its proximity to the segment. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table 3: Change in Exposure to Cholera at Boundary of Broad Street Pump Catchment Area

	Averaged Conditional Treatment Effects					
	Number of Deaths in Household (1)	House has at Least one Death (2)	Proportion of Deaths to Houses on Block (3)	Percent of Houses hit by Cholera on Block (4)	Proportion of Deaths to Houses in Neighborhood (5)	Percent of Houses hit by Cholera in Neighborhood (6)
Inside BSP area	0.231**	0.123**	0.160**	0.082***	0.232***	0.104***
	(0.101)	(0.051)	(0.064)	(0.026)	(0.045)	(0.019)
Boundary points	47	47	47	47	47	47
Bandwidth (meters)	57	57	57	57	57	57
Mean outside BSP area	0.241	0.139	0.248	0.142	0.255	0.146

Notes: In Columns (3) and (4), we define a neighborhood to include all houses on the block of the respective house, in addition to all of the houses on surrounding adjacent blocks. Optimal bandwidth determined as in Imbens and Kalyanaraman (2012) using a triangular kernel. The minimum of all optimal bandwidths for each boundary point is used. Since sample size used in analysis at each boundary point varies, no specific number of observations is reported. Boundary points are chosen randomly along the boundary and indicate the number of points at which the treatment effect is evaluated. Analysis is restricted to boundary points with at least one observation on each side of the BSP boundary within the optimal bandwidth (Boundary Positivity Assumption). Clustered standard errors by street block are shown in parenthesis. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table 4: Boundary Effects on Rental Prices, 1864

	Change in Log Rental Price, 1853-1864		Log Rental Price, 1864				
	Boundary RD	One-dimensional RD	Boundary RD	One-dimensional RD			
	Averaged Conditional Treatment Effects (1)	Optimal Band Parametric Form (2)	Averaged Conditional Treatment Effects (3)	Optimal Band Parametric Form (4)	Narrow Band Parametric Form (5)	Cluster by Street (6)	Infrastructure Only (7)
Inside BSP area	-0.112*	-0.138*	-0.139**	-0.149**	-0.110*	-0.149*	-0.119*
	(0.0639)	(0.0712)	(0.0653)	(0.0662)	(0.0633)	(0.0844)	(0.0690)
Distance to BSP boundary		-2.199		-3.000**	-1.582	-0.453*	-3.000**
		(2.408)		(1.384)		(3.712)	(1.427)
(Distance to BSP boundary)^2		10.81		11.90**	1.153	0.455*	11.90*
		(16.83)		(5.989)	(31.82)	(0.240)	(6.452)
(Distance to BSP boundary)^3		-18.43		-13.50*	9.343	-0.0723	-13.50
		(35.43)		(7.575)		(81.49)	(8.381)
Observations		491		737	415	737	743
Boundary points	47	.	47	.	.	.	.
Mean outside BSP area	0.0234	0.0252	54.74	55.42	48.48	55.42	55.42
Bandwidth (meters)	57	29.60	57	50.57	24	50.57	50.57

Notes: Clustered standard errors shown in parenthesis. Standard errors in Columns (1) and (3) obtained via bootstrap with 1,000 replications. Columns (2), (4), (5) and (7) use street blocks as clusters. Column (6) uses street clusters. Optimal bandwidth determined as in Imbens and Kalyanaraman (2012). Analysis in Columns (1) and (3) uses minimum of all optimal bandwidths for each boundary point. Since sample size used in analysis at each boundary point varies, Columns (1) and (3) do not report a specific number of observations. Boundary points are chosen randomly along the boundary and indicate the number of points at which the treatment effect is evaluated. Analysis is restricted to boundary points with at least one observation on each side of the BSP boundary within the optimal bandwidth (Boundary Positivity Assumption). Narrow bandwidth is defined as the smallest bandwidth for which the RD coefficient is statistically significant. Wide bandwidth is defined as the largest bandwidth for which the RD coefficient remains significant. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table 5: Boundary Effects on Residential Mobility

	Resident is Different in 1864 than in 1853				Log Rental Price, 1864	
	(1)	(2)	(3)	(4)	(5)	(6)
Inside BSP area	0.0728	0.110**	0.103**	0.00972	-0.118*	0.00575
	(0.0464)	(0.0422)	(0.0438)	(0.0594)	(0.0692)	(0.0631)
House has at least one death			0.0712*	0.0468	-0.0983**	-0.0618
			(0.0410)	(0.0406)	(0.0472)	(0.0418)
Number of deaths within neighborhood				0.00274**		-0.00426*
				(0.00132)		(0.00232)
Total number of houses in neighborhood				-0.00232**		0.00441**
				(0.00102)		(0.00198)
Distance to BSP boundary	-0.567	1.563	1.559	1.455	-3.090**	-3.410**
	(1.689)	(1.185)	(1.180)	(1.075)	(1.459)	(1.482)
(Distance to BSP boundary)^2	4.325	-7.223	-7.203	-6.934*	12.84*	14.40**
	(9.191)	(4.630)	(4.598)	(4.112)	(6.684)	(6.704)
(Distance to BSP boundary)^3	-9.338	7.627	7.626	7.296*	-15.08*	-16.46*
	(14.10)	(4.746)	(4.711)	(4.168)	(8.773)	(8.658)
Observations	765	1,056	1,031	1,031	724	724
Mean outside BSP area	0.576	0.656	0.656	0.656	55.42	55.42
Bandwidth (meters)	0.437	0.520	0.520	0.520	0.506	0.506

Notes: Clustered standard errors by street block shown in parenthesis. We define a neighborhood to include all of the houses on the block of the respective house, in addition to all of the houses on surrounding adjacent blocks. Columns (1), (5), and (6) use the optimal bandwidth as determined in Imbens and Kalyanaraman (2012) using a triangular kernel. Columns (2), (3), and (4) use the wide bandwidth defined as the largest bandwidth for which the RD coefficient remains statistically significant. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table 6: Boundary Effects on Migration Patterns by Cholera Exposure

	All of St. James		Broad Street Pump Catchment Area	
	(1) Resident is Different in 1864 than in 1853	(2) Resident is Different in 1864 than in 1853	(3) Resident is Different in 1864 than in 1853	(4) Resident is Different in 1864 than in 1853
House has at least one death	0.142**	0.142**	0.255***	0.317***
	(0.0603)	(0.0665)	(0.0820)	(0.0907)
Number of deaths within neighborhood	0.00371***		0.00352***	
	(0.00105)		(0.000979)	
(Household death)*(neighborhood deaths)	-0.00284**		-0.00519***	
	(0.00130)		(0.00151)	
Number of houses with death in neighborhood		0.00900***		0.00911***
		(0.00238)		(0.00283)
(Household death)*(neighborhood houses with death)		-0.00618*		-0.0149***
		(0.00329)		(0.00396)
Total number of houses in neighborhood	-0.00167	-0.00197	-0.000124	-0.000134
	(0.00123)	(0.00126)	(0.00126)	(0.00140)
Number of deaths in household	-0.00360	-0.00424	0.0151	0.0148
	(0.0176)	(0.0178)	(0.0196)	(0.0198)
Log total sums Assessed, 1853	-0.0375	-0.0326	-0.145**	-0.146**
	(0.0367)	(0.0367)	(0.0731)	(0.0722)
Observations	1,698	1,698	491	491
Mean among HHs with no deaths	0.521	0.521	0.602	0.602

Notes: This table shows the marginal effects of a probit estimation, with clustered standard errors by block shown in parenthesis. The number of deaths within the neighborhood and the number of houses in the neighborhood visited by cholera exclude the current household. In Columns (1) and (2), the mean outcome is reported among households with no deaths both within house and within neighborhood. In Columns (3) and (4), mean outcome is reported among households with no deaths within house and less than 5 deaths within their neighborhood - a lower bound of neighborhood deaths within the BSP subsample. We define a neighborhood to include all of the houses on the block of the respective house, in addition to all of the houses on surrounding adjacent blocks. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table 7: Boundary Effects on House Occupancy Characteristics

	Optimal Bandwidth			Narrow Bandwidth		
	(1)	(2)	(3)	(4)	(5)	(6)
	Number of occupants at address	Number of immigrant families at address	Proportion of immigrant families at address	Number of occupants at address	Number of immigrant families at address	Proportion of immigrant families at address
CENSUS DATA: 1851						
Inside BSP area	1.610	0.109	0.0159	1.840	0.107	0.0196
	(1.297)	(0.106)	(0.0264)	(1.339)	(0.111)	(0.0284)
Distance to BSP boundary	-0.385	0.00913	0.00199	-0.475	0.0419	-0.00129
	(0.577)	(0.0452)	(0.0121)	(0.825)	(0.0615)	(0.0173)
(Distance to BSP boundary)^2	0.0238	-0.00116	-0.000138	0.0308	-0.00409	0.000333
	(0.0431)	(0.00334)	(0.000880)	(0.0802)	(0.00591)	(0.00158)
(Distance to BSP boundary)^3	-0.000464	3.06e-05	3.32e-06	-0.000608	9.68e-05	-1.37e-05
	(0.000925)	(6.93e-05)	(1.86e-05)	(0.00226)	(0.000156)	(4.12e-05)
Observations	547	547	547	423	423	423
Bandwidth (meters)	31	31	31	24	24	24
Mean Outside BSP Area	13.569	0.413	0.119	13.628	0.393	0.110
CENSUS DATA: 1861						
Inside BSP area	4.707**	0.281**	0.0355	5.192**	0.318***	0.0555**
	(1.785)	(0.107)	(0.0240)	(1.975)	(0.110)	(0.0228)
Distance to BSP boundary	-0.425	0.00682	0.0103	-0.552	-0.0751	-0.00703
	(0.500)	(0.0548)	(0.0135)	(0.704)	(0.0843)	(0.0168)
(Distance to BSP boundary)^2	0.0339	0.000228	-0.000548	0.0489	0.00939	0.00139
	(0.0404)	(0.00378)	(0.00104)	(0.0682)	(0.00749)	(0.00166)
(Distance to BSP boundary)^3	-0.000707	-1.39e-05	1.00e-05	-0.00118	-0.000288	-4.89e-05
	(0.000890)	(7.75e-05)	(2.31e-05)	(0.00185)	(0.000194)	(4.52e-05)
Observations	483	483	483	377	377	377
Bandwidth (meters)	31	31	31	24	24	24
Mean Outside BSP Area	12.560	0.399	0.127	12.207	0.374	0.109

Notes: Clustered standard errors shown in parenthesis. Regressions include controls for proximity to latrines and sewage. Immigrant families are defined by a household with a head born outside of England, Wales, or Scotland. Optimal bandwidth determined as in Imbens and Kalyanaraman (2012) using a triangular kernel. Narrow bandwidth is defined as the smallest bandwidth for which the RD coefficient is statistically significant. Wide bandwidth is defined as the largest bandwidth for which the RD coefficient remains significant. Census data acquired from The National Archives of the UK: Public Record Office. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table 8: Boundary Effects on House Socioeconomic Status (1899)

	Very Poor		Poor		Working Poor		Middle Class	
	Averaged CTEs (1)	Optimal Band (2)	Averaged CTEs (3)	Optimal Band (4)	Averaged CTEs (5)	Optimal Band (6)	Averaged CTEs (7)	Optimal Band (8)
Inside BSP area	0.124**	0.086*	0.010	0.027	-0.031	-0.051	-0.103**	-0.072**
	(0.054)	(0.046)	(0.033)	(0.051)	(0.058)	(0.064)	(0.049)	(0.036)
Distance to BSP boundary		0.001		0.066		0.658		-1.438
		(0.747)		(1.326)		(2.086)		(1.583)
(Distance to BSP boundary)^2		-0.302		-1.450		0.177		8.822
		(2.176)		(7.317)		(13.512)		(10.846)
(Distance to BSP boundary)^3		0.218		2.853		-3.805		-17.151
		(1.929)		(11.340)		(23.952)		(20.640)
Observations	.	1,138	.	774	.	722	.	670
Boundary points	23	.	23	.	23	.	23	.
Bandwidth (meters)	0.93	76.1	0.93	43.2	0.93	38.2	0.93	34.9
Mean outside BSP Area	0.195	0.207	0.233	0.236	0.416	0.384	0.156	0.150

Notes: Clustered standard errors shown in parenthesis. Standard errors in Columns (1), (3), (5), and (7) obtained via bootstrap with 500 replications. Columns (2), (4), (6) use street blocks as clusters. Optimal bandwidth determined as in Imbens and Kalyanaraman (2012). Analysis in Columns (1), (3), and (5) uses minimum of all optimal bandwidths for each boundary point. Since sample size used in analysis at each boundary point varies, Columns (1), (3), (5), and (7) do not report a specific number of observations. Boundary points are chosen randomly along the boundary and indicate the number of points at which the treatment effect is evaluated. Analysis restricted to boundary points with at least one observation on each side of the BSP boundary within the optimal bandwidth (Boundary Positivity Assumption). All specification include controls for distances to Soho centroid, public squares, theaters, police stations, primary schools, pubs, churches, and banks. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table 9: Boundary Effects on Rental Prices, 1936

	Log Rental Price, 1936				
	Boundary RD	One-dimensional RD			
	Averaged Conditional Treatment Effects (1)	Optimal Band Parametric Form (2)	Narrow Band Parametric Form (3)	Optimal Band Cluster by Street (4)	Optimal Band Segment FE (5)
Inside BSP area	-0.346*	-0.366**	-0.301*	-0.366**	-0.356**
	(0.195)	(0.152)	(0.175)	(0.125)	(0.167)
Distance to BSP boundary		-2.362	-8.098	-2.362	-2.975
		(5.323)	(11.128)	(4.049)	(5.657)
(Distance to BSP boundary)^2		6.020	64.594	6.020	7.167
		(29.531)	(84.175)	(22.094)	(32.117)
(Distance to BSP boundary)^3		-1.745	-158.127	-1.745	-2.488
		(47.328)	(183.920)	(36.583)	(51.402)
Observations	.	230	180	230	230
Boundary points	25	.	.	.	.
Mean outside BSP area (in £)	448.12	459.22	440.18	459.22	459.22
Ave. monthly rental price (in 2014 £)	2397.22	2,371.81	2,306.17	2,371.81	2,371.81
Clusters	.	51	39	22	51
Bandwidth (meters)	70	39.85	29	39.85	39.85

Notes: Clustered standard errors shown in parenthesis. Standard errors in Column (1) are obtained via bootstrap with 1,000 replications. Columns (2), (3), and (5) use blocks as clusters. Column (4) uses street clusters. Optimal bandwidth determined as in Imbens and Kalyanaraman (2012). Analysis in Column (1) uses minimum of all optimal bandwidths for each boundary point. Since sample size used in analysis at each boundary point varies, Column (1) does not report a specific number of observations. Boundary points are chosen randomly along the boundary and indicate the number of points at which the treatment effect is evaluated. Analysis is restricted to boundary points with at least one observation on each side of the BSP boundary within the optimal bandwidth (Boundary Positivity Assumption). Narrow bandwidth is defined as the smallest bandwidth for which the RD coefficient is statistically significant. Segments in Column (5) are determined by dividing the BSP boundary into five segments of equal length respectively. An observation is assigned to a segment based on its proximity to the segment. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table 10: Boundary Effects on House Prices, Zoopla House Value Estimates, and Rental Prices 1995-2013, 2015

	House Prices and Zoopla Estimates			Zoopla Estimates Only			House Prices Only		House Rental Prices	
	Averaged CTEs (1)	Optimal Band (2)	Narrow Band (3)	Averaged CTEs (4)	Optimal Band (5)	Narrow Band (6)	Optimal Band (7)	Narrow Band (8)	Optimal Band (9)	Narrow Band (10)
Inside BSP area	-0.158**	-0.238**	-0.231***	-0.124**	-0.168**	-0.199**	-0.286**	-0.302**	-0.163**	-0.123*
	(0.069)	(0.095)	(0.079)	(0.054)	(0.073)	(0.090)	(0.128)	(0.121)	(0.076)	(0.068)
Distance to BSP boundary		3.610	10.665**		11.229***	5.585	-3.674	-5.698	1.446	-2.566
		(3.802)	(4.439)		(4.232)	(4.601)	(3.361)	(8.489)	(2.713)	(3.209)
Distance to BSP boundary^2		-19.227	-73.187**		-85.607***	-29.055	28.713	41.640	-0.001	0.001
		(22.110)	(29.806)		(28.159)	(35.783)	(17.744)	(56.658)	(0.001)	(0.001)
Distance to BSP boundary^3		29.066	133.860**		166.638***	28.963	-48.880*	-71.147	0.000	-0.000
		(34.124)	(53.740)		(52.307)	(77.702)	(27.656)	(99.511)	(0.000)	(0.000)
Observations	.	717	679	.	466	311	221	211	176	141
Boundary points	26	.	.	26	.	.	.	.	.	.
Mean outside BSP area (in 2014 £)	963.83	930.92	931.81	963.84	945.28	967.82	902.88	904.53	0.752	0.752
Clusters	.	218	198	.	191	169	30	27	53	50
Bandwidth (meters)	85.0	41.2	37	85.0	35.1	30	43.5	37	50	44.5

Notes: Clustered standard errors shown in parenthesis. Standard errors in Columns (1) and (3) obtained via bootstrap with 1,000 replications. Columns (2), (3), (5)-(10) use postal codes as clusters. Optimal bandwidth determined as in Imbens and Kalyanaraman (2012). Analysis in Columns (1) and (3) uses minimum of all optimal bandwidths for each boundary point. Since sample size used in analysis at each boundary point varies, Columns (1) and (3) do not report a specific number of observations. Boundary points are chosen randomly along the boundary and indicate the number of points at which the treatment effect is evaluated. Analysis is restricted to boundary points with at least one observation on each side of the BSP boundary within the optimal bandwidth (Boundary Positivity Assumption). Narrow bandwidth is defined as the smallest bandwidth for which the RD coefficient is statistically significant. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table 11: Boundary Effects using John Snow's boundary definition

	Pre-outbreak	Cholera Exposure		Post-outbreak		
	Rental price 1853 (1)	Number of Deaths in Household (2)	House has at least one death (3)	Rental price 1864 (4)	Rental price 1936 (5)	House value 1995-2013, 2015 (6)
Inside BSP	-0.054	0.488***	0.272***	-0.127*	-0.277*	-0.226**
	(0.069)	(0.089)	(0.031)	(0.068)	(0.154)	(0.090)
Observations	777	818	818	754	208	756
R-squared	0.464	0.067	0.100	0.379	0.506	0.384
Clusters	104	94	94	104	45	224
Bandwidth (meters)	55.6	47.6	47.6	55.8	38.6	41.0

Notes: Bandwidth chosen optimally as in Imbens and Kalyanaraman (2012). *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table 12: False Treatment Boundary Tests

	Pre-Outbreak (1853)	Cholera Exposure		Post-outbreak (1864)			Current (1995-2013, 2015)	
	Rental price	Number of Deaths in Household	House has at least one death	Rental price	Change in Rental price	Different Resident in 1864	Sales price	Price and Zoopla estimates
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. False Boundary 1								
RD coeffcient	0.132	-0.009	0.016	0.049	0.035	0.177		
	(0.173)	(0.036)	(0.053)	(0.178)	(0.058)	(0.163)		
Observations	153	274	236	129	139	224		
Bandwidth	0.440	0.728	0.619	0.370	0.401	0.584		
Panel B. False Boundary 2								
RD coeffcient	-0.035	-0.254	0.012	-0.166	-0.115	-0.133	0.045	-0.181
	(0.162)	(0.204)	(0.073)	(0.141)	(0.069)	(0.145)	(0.159)	(0.150)
Observations	227	266	267	303	249	308	190	455
Bandwidth	0.497	0.515	0.520	0.663	0.565	0.584	0.550	0.550
Panel C. False Boundary 3								
RD coeffcient	0.038	0.069	-0.025	0.055	0.053	-0.180*		
	(0.115)	(0.084)	(0.036)	(0.054)	(0.127)	(0.105)		
Observations	230	507	567	341	249	649		
Bandwidth	0.300	0.646	0.744	0.462	0.338	0.993		
Panel D. False Boundary 4								
RD coeffcient							-0.050	-0.069
							(0.199)	(0.125)
Observations							217	453
clusters							31	90
Bandwidth							0.551	0.551

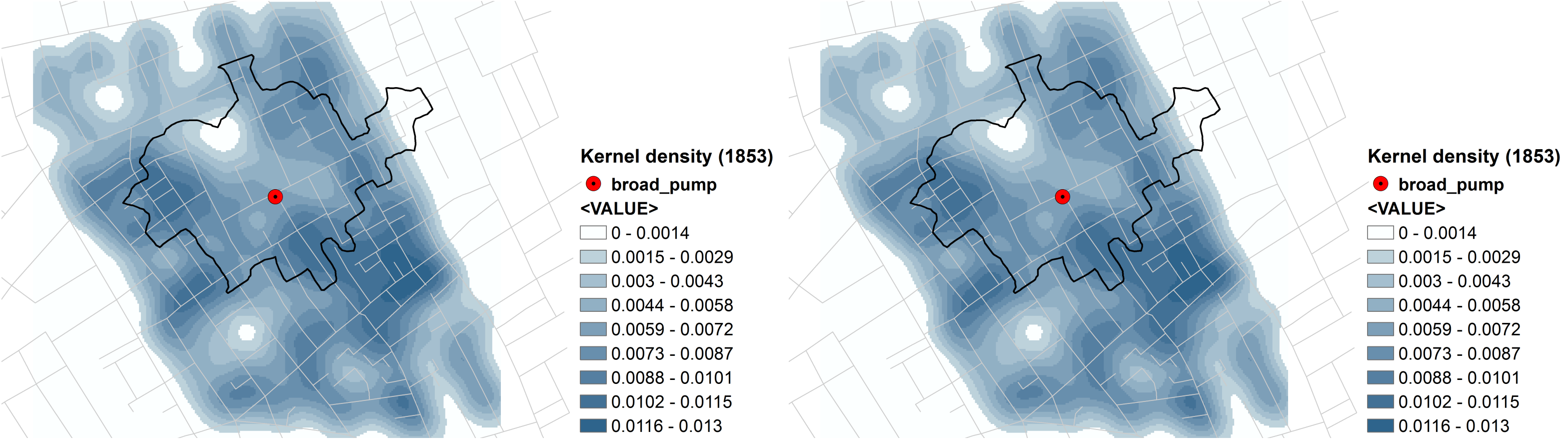
Notes: Clustered standard errors in parentheses. Columns (1)-(6) use street blocks as clusters. Columns (7) and (8) use post codes. Optimal bandwidth chosen using Imbens and Kalyanaraman (2012)

Table 13: Rental Price, House Occupancy and Cholera Exposure within BSP Area

	(1) Log Rental Price	(2) Number of Immigrant Families at Address	(3) Proportion of Immigrant Families at Address
Panel A. Pre-Outbreak (1851,1853)			
House had at least one death (1854)	-0.069	0.080	0.010
	(0.045)	(0.085)	(0.022)
Observations	457	458	458
Mean (houses with no deaths)	48.76	0.407	0.105
Panel B. Post-Outbreak (1861,1864)			
House had at least one death (1854)	-0.050	0.186	0.025
	(0.043)	(0.132)	(0.042)
Observations	447	414	414
Mean (houses with no deaths)	50.35	0.665	0.162

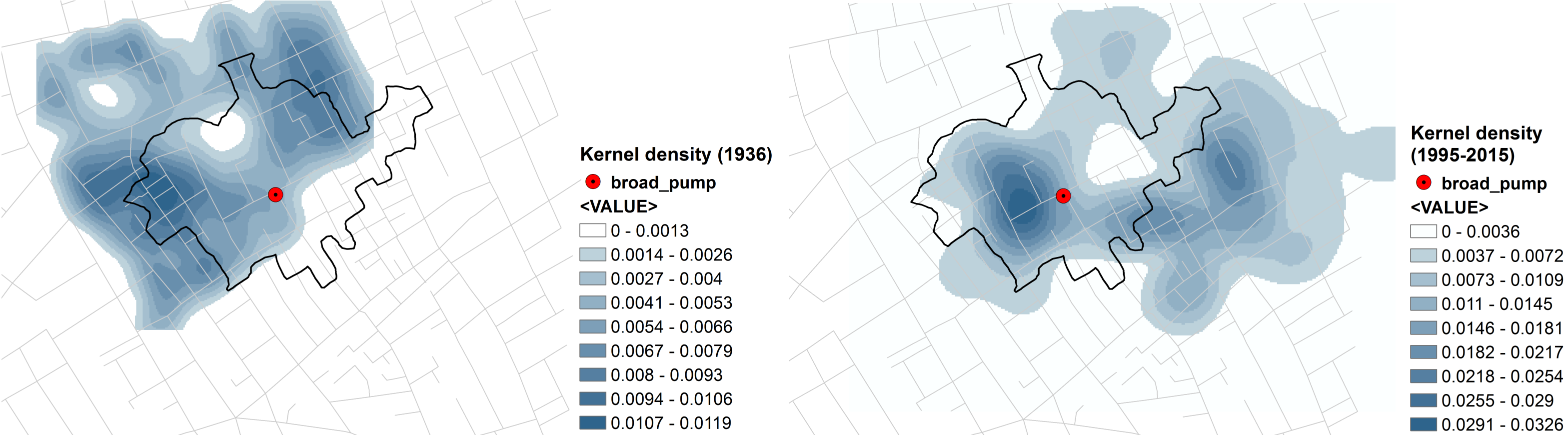
Notes: Clustered standard errors shown in parenthesis. All regressions restricted to properties inside the BSP boundary. Column (1) uses Land Tax records. Columns (2)-(4) use Census records. Column (1) gives results for the year 1853 and 1864 for the pre-outbreak and post-outbreak period, respectively. Columns (2)-(4) give results for the year 1851 and 1861 for the pre-outbreak and post-outbreak period, respectively. Immigrant families are defined by a household with a head born outside of England, Wales, or Scotland. Analysis uses block fixed effects. Optimal bandwidth determined as in Imbens and Kalyanaraman (2012) using a triangular kernel. Census data acquired from The National Archives of the UK: Public Record Office. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Appendix B: Additional Figures and Tables



(a) 1853

(b) 1864



(c) 1936

(d) 1995-2013, 2015

Figure B1: Kernel Density Estimates

Notes: Kernel estimates of the density of properties around the BSP boundary using 50-meter bandwidth. Obtained using the Kernel Density tool in ArcGIS Spatial Analyst package.

Table B1: House characteristics (1853) with Conley (1999) Standard errors

	Full sample			Within 100 m			Optimal bandwidth	
	Inside (1)	Outside (2)	S.E. (3)	Inside (4)	Outside (5)	S.E. (6)	RD (7)	S.E. (8)
Rental characteristics:								
Rental price (in logs)	3.723	3.780	(0.059) [0.052]	3.719	3.743	(0.067) [0.058]	0.126	(0.113) [0.112]
Tax assessed (in logs)	0.456	0.518	(0.058) [0.051]	0.453	0.496	(0.065) [0.056]	0.110	(0.114) [0.112]
Tax exonerated (yes=1)	0.067	0.230	(0.040)*** [0.037]***	0.067	0.22	(0.052)*** [0.048]***	-0.005	(0.059) [0.057]
Sewer access:								
Old/Existing	0.477	0.564	(0.085) [0.074]	0.473	0.589	(0.091) [0.079]	0.082	(0.135) [0.135]
New sewer	0.396	0.276	(0.081) [0.072]*	0.399	0.260	(0.084) [0.075]*	0.116	(0.105) [0.108]
No access	0.127	0.159	(0.055) [0.047]	0.128	0.151	(0.063) [0.052]	-0.198	(0.100)* [0.100]*
Distance (m/100) to:								
Closest pump	1.046	0.957	(0.079) [0.068]	1.052	1.066	(0.093) [0.079]	0.111	(0.079) [0.078]
Soho centroid	1.320	2.471	(0.119)*** [0.101]***	1.325	2.177	(0.131)*** [0.114]***	-0.104	(0.202) [0.199]
Presumed plague pit	2.359	3.135	(0.224)*** [0.181]***	2.358	2.630	(0.215) [0.184]	0.260	(0.301) [0.296]
Public square	2.586	2.717	(0.135) [0.114]	2.584	2.715	(0.138) [0.121]	-0.109	(0.189) [0.194]
Church	1.322	1.712	(0.129)*** [0.112]***	1.323	1.609	(0.141)** [0.124]**	0.071	(0.163) [0.168]
Police station	4.361	5.413	(0.261)*** [0.210]***	4.364	4.797	(0.221)* [0.191]**	-0.078	(0.438) [0.420]
Fire engine	3.603	2.664	(0.187)*** [0.163]***	3.597	2.864	(0.227)*** [0.196]***	0.452	(0.353) [0.339]
Theater	4.004	5.306	(0.232)*** [0.189]***	4.011	4.679	(0.219)*** [0.189]***	-0.397	(0.335) [0.327]
Pub	0.286	0.408	(0.037)*** [0.032]***	0.287	0.406	(0.046)** [0.038]***	-0.184	(0.111) [0.105]*
Urinal	0.878	1.122	(0.087)*** [0.075]***	0.874	1.019	(0.088) [0.080]*	-0.144	(0.141) [0.143]
Sewer ventilator	0.429	0.555	(0.048)*** [0.042]***	0.431	0.563	(0.050)*** [0.045]***	-0.023	(0.100) [0.098]
Primary school	1.306	2.474	(0.132)*** [0.110]***	1.306	2.023	(0.109)*** [0.099]***	-0.152	(0.195) [0.196]
Bank	3.949	4.694	(0.316)** [0.255]***	3.947	4.095	(0.315) [0.268]	-0.005	(0.520) [0.502]
Observations	495	1230		491	815		534	

Notes: Columns (1), (2), (4), and (5) give the mean of the corresponding variable. Columns (3) and(6) give the clustered standard error at the street block level in parenthesis and Conley (1999) standard errors in brackets for the difference in means. Conley (1999) standard errors use a distance cutoff equal to the average length of a street block and a Bartlett spatial weighting kernel. “Inside” and “Outside” indicate whether a house is inside or outside the BSP area respectively. Columns (7) and (8) give the estimated coefficient and standard errors for the RD specification that uses the corresponding variable as its outcome. Optimal bandwidth is determined as in Imbens and Kalyanaraman (2012) using a triangular kernel. Column (7) uses the minimum of all optimal bandwidths (27.50 meters). *, **, and *** indicate 10, 5, and 1 percent significance respectively.

Table B2: Change in Exposure to Cholera at Boundary of Broad Street~~.~~ Pump Catchment Area

	(1) Number of deaths in household	(2) House has at least one death	(3) Proportion of deaths to houses on block	(4) Percent of houses hit by cholera on block	(5) Proportion of deaths to houses in neighborhood	(6) Percent of houses hit by cholera in neighborhood
Inside BSP area	0.548***	0.238***	0.427***	0.191***	0.459***	0.210***
	(0.0939)	(0.0404)	(0.0885)	(0.0363)	(0.0655)	(0.0297)
Distance to BSP boundary	1.472	0.890	-0.969	-0.218	0.152	0.390
	(2.241)	(0.979)	(2.361)	(0.693)	(1.101)	(0.429)
(Distance to BSP boundary)^2	-2.151	-2.972	6.618	1.140	2.711	-0.762
	(8.982)	(4.065)	(10.44)	(3.493)	(4.959)	(1.984)
(Distance to BSP boundary)^3	-0.682	2.337	-10.67	-1.773	-5.577	0.320
	(10.45)	(4.791)	(13.59)	(5.093)	(6.189)	(2.518)
Observations	949	929	825	792	885	922
Bandwidth (meters)	57.55	55.54	48.38	45.83	50.10	52.90
Mean outside BSP area	0.239	0.141	0.286	0.165	0.267	0.149

Notes: In Columns (3) and (4), we define a neighborhood to include all of the houses on the block of the respective house, in addition to all of the houses on surrounding adjacent blocks. Optimal bandwidth determined as in Imbens and Kalyanaraman (2012) using a triangular kernel. Clustered standard errors by street block shown in parenthesis. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table B3: Change in Exposure to Cholera at Boundary of Broad Street~~.~~ Pump Catchment Area, Narrow BW

	(1) Number of deaths in household	(2) House has at least one death	(3) Proportion of deaths to houses on block	(4) Percent of houses hit by cholera on block	(5) Proportion of deaths to houses in neighborhood	(6) Percent of houses hit by cholera in neighborhood
Inside BSP area	0.336***	0.140***	0.250***	0.113***	0.220***	0.107***
	(0.114)	(0.0483)	(0.0725)	(0.0294)	(0.0496)	(0.0247)
Distance to BSP boundary	-0.0142	-0.0254	-0.0554	-0.0187*	0.0217	0.00254
	(0.0991)	(0.0355)	(0.0445)	(0.0109)	(0.0194)	(0.00658)
(Distance to BSP boundary)^2	0.00411	0.00383	0.00592*	0.00210**	-0.00139	0.000138
	(0.00854)	(0.00314)	(0.00340)	(0.000945)	(0.00180)	(0.000627)
(Distance to BSP boundary)^3	-0.000166	-0.000129	-0.000177**	-6.60e-05**	2.29e-05	-1.15e-05
	(0.000213)	(8.00e-05)	(8.54e-05)	(2.59e-05)	(4.98e-05)	(1.79e-05)
Observations	462	462	473	473	473	473
Bandwidth (meters)	24	24	24	24	24	24
Mean outside BSP area	0.312	0.202	0.338	0.195	0.339	0.185

Notes: In Columns (3) and (4), we define a neighborhood to include all of the houses on the block of the respective house, in addition to all of the houses on surrounding adjacent blocks. Narrow bandwidth is defined as the smallest bandwidth for which the RD coefficient is statistically significant. Clustered standard errors by street block shown in parenthesis. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table B4: Change in Exposure to Cholera at Boundary of Broad Street~~.~~ Pump Catchment Area, Wide BW

	(1) Number of deaths in household	(2) House has at least one death	(3) Proportion of deaths to houses on block	(4) Percent of houses hit by cholera on block	(5) Proportion of deaths to houses in neighborhood	(6) Percent of houses hit by cholera in neighborhood
Inside BSP area	0.531***	0.242***	0.492***	0.222***	0.523***	0.220***
	(0.0944)	(0.0404)	(0.0905)	(0.0362)	(0.0692)	(0.0299)
Distance to BSP boundary	0.00737	0.0104	0.00616	0.00161	0.0145	0.00511
	(0.0232)	(0.00962)	(0.0205)	(0.00609)	(0.0105)	(0.00399)
(Distance to BSP boundary)^2	0.000157	-0.000377	-0.000180	-5.94e-05	-0.000408	-0.000135
	(0.000959)	(0.000397)	(0.000789)	(0.000248)	(0.000447)	(0.000174)
(Distance to BSP boundary)^3	-5.72e-06	3.44e-06	1.49e-06	3.65e-07	4.21e-06	1.13e-06
	(1.14e-05)	(4.66e-06)	(8.99e-06)	(2.87e-06)	(5.38e-06)	(2.12e-06)
Observations	933	933	956	956	956	956
R-squared	0.072	0.093	0.226	0.329	0.417	0.442
Bandwidth (meters)	56	56	56	56	56	56
Mean outside BSP area	0.244	0.141	0.250	0.143	0.257	0.146

Notes: In Columns (3) and (4), we define a neighborhood to include all of the houses on the block of the respective house, in addition to all of the houses on surrounding adjacent blocks. Wide bandwidth is defined as the largest bandwidth for which the RD coefficient remains significant. Clustered standard errors by street block shown in parenthesis. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table B5: Boundary Effects on Total Taxes Assessed, 1864

	Change in Taxes Assessed, 1853-1864		Log Total Taxes, 1864				
	(1) Optimal Band Parametric Form	(2) Optimal Band Local Linear Regression	(3) Optimal Band Parametric Form	(4) Narrow Band Parametric Form	(5) Wide Band Parametric Form	(6) Optimal Band Cluster by Street	(7) Optimal Band Infrastructure Only
Inside BSP area	-0.129*	-0.166*	-0.131**	-0.115*	-0.0720	-0.131*	-0.139*
	(0.0701)	(0.0949)	(0.0608)	(0.0623)	(0.0660)	(0.0779)	(0.0724)
Distance to BSP boundary	-1.637	-0.610	-4.143**	-0.449*	-1.220	-4.143**	-2.341
	(2.219)	(0.411)	(1.874)	(0.229)	(0.762)	(1.901)	(1.844)
(Inside BSP)*(Distance to boundary)		0.00432					
		(0.00431)					
(Distance to BSP boundary)^2	5.609		19.08*	0.497**	2.726	19.08*	7.407
	(14.79)		(10.49)	(0.231)	(2.272)	(10.25)	(10.80)
(Distance to BSP boundary)^3	-5.959		-25.97	-0.0821	-1.811	-25.97	-5.415
	(29.63)		(16.87)	(0.0561)	(1.889)	(15.95)	(18.30)
Observations	568	568	689	1,662	1,123	689	693
Mean outside BSP area	0.316	0.316	0.395	24	0.395	0.395	0.395
Bandwidth (meters)	19.9	19.9	212.5	187.2	212.5	212.5	212.5

Notes: Clustered standard errors shown in parenthesis. Columns (1), (2), (3), (4), (5) and (7) use street blocks as clusters. Column (6) uses street clusters. Optimal bandwidth determined as in Imbens and Kalyanaraman (2012) using a triangular kernel. Narrow bandwidth is defined as the smallest bandwidth for which the RD coefficient is statistically significant. Wide bandwidth is defined as the largest bandwidth for which the RD coefficient remains significant. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table B6: Charles Booth class categorization

Class code	Description of class
A	The lowest class which consists of some occasional labourers, street sellers, loafers, criminals and semi-criminals. Their life is the life of savages, with vicissitudes of extreme hardship and their only luxury is drink.
B	Casual earnings, very poor. The labourers do not get as much as three days work a week, but it is doubtful if many could or would work full time for long together if they had the opportunity. Class B is not one in which men are born and live and die so much as a deposit of those who from mental, moral and physical reasons are incapable of better work.
C	Intermittent earning. 18s to 21s per week for a moderate family. The victims of competition and on them falls with particular severity the weight of recurrent depressions of trade. Labourers, poorer artisans and street sellers. This irregularity of employment may show itself in the week or in the year: stevedores and waterside porters may secure only one of two days' work in a week, whereas labourers in the building trades may get only eight or nine months in a year.
D	Small regular earnings. poor, regular earnings. Factory, dock, and warehouse labourers, carmen, messengers and porters. Of the whole section none can be said to rise above poverty, nor are many to be classed as very poor. As a general rule they have a hard struggle to make ends meet, but they are, as a body, decent steady men, paying their way and bringing up their children respectably
E	Regular standard earnings, 22s to 30s per week for regular work, fairly comfortable. As a rule the wives do not work, but the children do: the boys commonly following the father, the girls taking local trades or going out to service.
F	Higher class labour and the best paid of the artisans. Earnings exceed 30s per week. Foremen are included, city warehousemen of the better class and first hand lightermen; they are usually paid for responsibility and are men of good character and much intelligence.
G	Lower middle class. Shopkeepers and small employers, clerks and subordinate professional men. A hardworking sober, energetic class.
H	Upper middle class, servant keeping class.

Table B7: Boundary Effect on Total Taxes Assessed, 1936

	(1) Optimal band Parametric Form	(2) Narrow band Parametric Form	(3) Wide band Parametric Form	(4) Optimal band Cluster by Street	(5) Optimal band Segment FE
Inside BSP area	-0.198*	-0.195*	-0.437***	-0.198	-0.162
	(0.117)	(0.115)	(0.135)	(0.130)	(0.131)
Distance to BSP boundary	-2.167	-1.678	-1.208	-2.167	-2.765
	(2.578)	(2.781)	(1.285)	(2.362)	(2.387)
(Distance to BSP boundary)^2	8.081	5.381	4.120	8.081	11.633
	(11.762)	(13.748)	(3.192)	(10.337)	(11.116)
(Distance to BSP boundary)^3	-6.282	-2.158	-2.916	-6.282	-11.527
	(15.834)	(19.746)	(2.302)	(13.607)	(15.037)
Observations	450	427	699	450	450
Mean outside BSP area	2.123	2.128	2.153	2.123	2.123
Clusters	62	61	83	26	62
Bandwidth (meters)	48.4	45.0	96.8	48.4	48.4

Notes: Clustered standard errors shown in parenthesis. Columns (1), (2), (3), and (5) use street blocks as clusters. Column (4) uses streets. Optimal bandwidth determined as in Imbens and Kalyanaraman (2012) using a triangular kernel. Narrow bandwidth is defined as the smallest bandwidth for which the RD coefficient is statistically significant. Wide bandwidth is defined as the largest bandwidth for which the RD coefficient remains significant. Segments in Column (5) are determined by dividing the BSP boundary into five segments of equal length respectively. An observation is assigned to a segment based on its proximity to the segment. *, **, *** indicate 10, 5, and 1 percent significance respectively.

Table B8: Boundary Effects on House Prices Only, 1995-2013, 2015

	(1) Optimal band Parametric Form	(2) Narrow band Parametric Form	(3) Wide band Parametric Form	(4) Optimal band Cluster by Street	(5) Wide band Segment FE
Inside BSP area	-0.318**	-0.302**	-0.170	-0.286	-0.066
	(0.133)	(0.121)	(0.162)	(0.169)	(0.220)
Distance to BSP boundary	0.335	-5.698	3.426*	-3.674	4.996***
	(5.273)	(8.489)	(1.722)	(3.971)	(1.720)
(Distance to BSP boundary)^2	7.462	41.640	-9.582*	28.713	-14.723***
	(28.188)	(56.658)	(5.681)	(21.645)	(5.458)
(Distance to BSP boundary)^3	-20.224	-71.147	7.181	-48.880	11.401**
	(42.474)	(99.511)	(4.821)	(33.996)	(4.440)
Observations	221	211	490	221	490
Mean outside BSP area	902,881	904,533	1.019e+06	902,881	1.019e+06
Segments	.	.	.	.	4
Clusters	30	27	64	16	64
Bandwidth (meters)	0.435	0.370	0.869	0.435	0.869

Notes: Clustered standard errors shown in parenthesis. Columns (1), (2), (3), and (5) use postal codes as clusters. Column (4) uses streets as clusters. Optimal bandwidth determined as in Imbens and Kalyanaraman (2012) using a triangular kernel. Narrow bandwidth is defined as the smallest bandwidth for which the RD coefficient is statistically significant. Wide bandwidth is defined as the largest bandwidth for which the RD coefficient remains significant. Clusters in Column (4) and segments in (5) are determined by dividing the BSP boundary into ten and five segments of equal length respectively. An observation is assigned to a segment based on its proximity to the segment. *, **, *** indicate 10, 5, and 1 percent significance respectively.