American Economic Review 2020, 110(2): 475–525
https://doi.org/10.1257/aer.20190759

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

By Attila Ambrus, Erica Field, and Robert Gonzalez*

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 one neighborhood of nineteenth century London. Ten years after the epidemic, housing prices are significantly lower just inside the catchment area of the water pump that transmitted the disease. Moreover, differences in housing prices persist over the following 160 years. We make sense of these patterns by building a model of a rental market with frictions in which poor tenants exert a negative externality on their neighbors. This showcases how a locally concentrated income shock can persistently change the tenant composition of a block. (*JEL* D62, O18, R21, R31)

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*

Can disease 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 which 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 in a city. Yet, in reality, spatial discontinuities in urban land values are frequently observed and do not always appear related to discrete changes in local amenities.

We examine this question in the context of a cholera epidemic that hit a single urban parish of London in 1854. Over the course of one month, 660 residents living

*Ambrus: Department of Economics, Duke University (email: attila.ambrus@duke.edu); Field: Department of Economics, Duke University (email: field.erica@duke.edu); Gonzalez: Department of Economics, University of South Carolina (email: robert.gonzalez@moore.sc.edu). Hilary Hoynes was the coeditor for this article. We thank Patrick Bayer, Sarah Gelfand, and seminar participants at the University of Stockholm, Stockholm School of Economics, Northwestern, University of Michigan, SITE, Harvard/MIT, Wharton, Brown, University of Washington, LSE, Oxford, and University of Houston for useful feedback, and Volodymyr Baranovskyi, Giudy Rusconi, and Veronica Polin for excellent research assistance. All errors are our own.

†Go to https://doi.org/10.1257/aer.20190759 to visit the article page for additional materials and author disclosure statements.

in the 0.5-mile radius of St. James Parish died from cholera, implying that roughly 5 percent 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 13 wells serving the parish. As a result, the impact was concentrated in one particular neighborhood wherein a much higher fraction of families, 13 percent, experienced a loss.

In this paper, we test empirically whether the cholera outbreak had a causal effect on neighborhood poverty, as measured by real estate prices, soon and long after the epidemic ended. There are two main reasons for focusing on this particular event. First, the disease pattern of the 1854 outbreak 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 nineteenth century epidemiologist, John Snow, since disease exposure was largely contained to one standing water source, there are sharp changes in death rates 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 at work.[1]

Our results reveal that houses inside the catchment area suffer a roughly 15 percent loss in rental value within a decade of the epidemic. More surprisingly, differences in property values persist for 160 years and show no signs of convergence. In 1936, we estimate 37 percent lower property values just inside relative to just outside the catchment area. Differences remain large and significant in contemporary real estate sales prices. Associated with the fall in rental prices is a significant increase in residential crowding at the boundary that emerges shortly after the epidemic. This pattern is consistent with historic accounts of renters taking in short-term tenants when money was tight, which had the potential to produce immediate block-level externalities, particularly in an already crowded neighborhood such as St. James.

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 when poor tenants generate negative externalities on rich neighbors. A key component of the model is that a landlord can only change the composition of a block gradually over time as tenants' rental contracts expire idiosyncratically. In this setting, we investigate how the landlord's optimal strategy regarding the type of tenants to attract changes when multiple tenants on his block simultaneously experience an income shock. As we show in the paper, if the fraction of households on the block that simultaneously 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.

[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-twentieth century. Land tax records were not well preserved in many parts of the city.

Back-of-the-envelope calculations using historic data on returns to capital indicate that such transitions could occur under reasonable parameter values.

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 of all properties in 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 our model's predictions.

Our results tie into a large literature in economic geography on the 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.[2] Our paper advances this discussion by undertaking a similar exercise within the micro-environment 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 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 of 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 be just as important in explaining income divergence across and not just within local economies. 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 communities than has been considered in the literature.

[2] For instance, Bleakley and Lin (2012) shows how geographic features (portage) that contributed to economic activity historically but stopped mattering in the nineteenth century are correlated with the long-run income growth of cities. Similarly, Hanlon (2017) shows that a short-lived price shock to the cotton industry in nineteenth 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 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 nineteenth century.

Our findings also illustrate the potential economic cost of spatially correlated shocks when there are significant externalities from neighbors' socioeconomic 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 or other attempts to upgrade poor neighborhoods located on intrinsically valuable property. On the other hand, our results also imply that those affected by disease epidemics such as the recent cholera epidemic in Haiti or the Ebola outbreak in West Africa will be less likely to be priced out of their current neighborhood, which could serve as a form of targeted insurance to disease victims.

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.[3] First, while agents in the models above follow simple behavioral rules, 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.[4] On a technical level, our paper is related to asynchronous-move dynamic games, and the role of asynchronicity of moves in coordination problems.[5]

The remainder of this paper proceeds as follows. Section I describes the setting and data. Section II describes the empirical strategy. Section III examines the immediate and long-term impacts of the cholera outbreak on housing prices. Section IV provides a theoretical analysis of one channel of persistence, and other potential channels are considered in Section V. Section VI concludes.

I. Background

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

[3] Refer to Section IV for a more detailed description of the related work and our contribution.

[4] Möbius (2000) and Guerrieri, Hartley, and Hurst (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, Hartley, and Hurst (2013) is that poor agents exert negative externality on their neighbors. Hornbeck and Keniston (2017) also investigates housing choices with externalities, but in a very different context, in which the quality of neighboring buildings affects the incentives of a home-owner to invest in the quality of her house. This leads to very different long-term dynamics than in our paper. 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.

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

A. *The Broad Street Cholera Outbreak of 1854*

In 1854, St. James, what is now the neighborhood of Soho, was a working-class neighborhood of 35,000 residents and a heavy commercial district that housed a large number of self-employed.[6] 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 previously been a wealthy neighborhood that contained many multi-story buildings, and 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 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, p. 18).

The majority of occupants of St. James were renters (93 percent) with absentee landlords who owned multiple flats on the block.[7] Rental contracts 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). Long-term leases were also possible, as were "tenancy-at-will" agreements, in which either party could leave at any time. However, rental data from land tax records indicate that nearly all contracts were at least a year, consistent with historic accounts.[8] Thus, while tenants had weak rights by twentieth 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 13 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. 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 who lived within the catchment area of the Broad Street pump, which encompassed 57 densely packed blocks.[9] The epidemic was fast and furious, and, because the source of transmission was stagnant rather than circulating, highly

[6] Much has been written by historians about the character of this particularly colorful neighborhood. See, for instance, Summers (1991).

[7] Source: 1853 UK Land Tax Assessment.

[8] Anecdotally, 3-year and 5-year contracts were also common. 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).

[9] 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.

geographically concentrated. By the epidemic's close within a month, 660 residents of St. James had died from cholera, or 3 percent of the population. Within the Broad Street pump (BSP) catchment area, an estimated 16 percent of residents had contracted the disease and approximately 8 percent died.

One particular public health authority, Dr. John Snow, had postulated after studying patterns of past outbreaks that cholera was transmitted through water. In order to collect evidence to support his theory, he immediately began mapping victims of the outbreak alongside information on the location of wells within the parish, which revealed a stark pattern of disease incidence in which nearly all victims were clustered around Broad Street pump. 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 reinstalled the Broad Street pump. The epidemiological analysis conducted by 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.

B. *The Impact of Cholera on Neighborhood Poverty*

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.

Conceptually, we anticipate changes in the rental value of property in the neighborhood arising from the sudden impoverishment of a substantial number of its residents. Hence, it is first worth considering the scale of the epidemic within the Broad Street pump neighborhood, which contained approximately 500 properties and 2,000 families.[10] By our estimates, 42 percent of all properties in the neighborhood experienced at least one cholera death during the epidemic, with 19 percent of households losing just one member and 23 percent losing more than one. This implies that approximately 5 percent of families lost one member and another 9 percent of families lost at least two members.[11]

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 76 percent of deaths occurred among individuals aged 10 to 60 (Cholera Inquiry Committee 1855). This implies that, in expectation, 13 percent of families inside the Broad Street pump catchment area (and only 1 percent in the rest of St. James) lost a potential wage-earner. In other words, 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

[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 parish contains 1,650 addresses, 4,439 families, and 20,807 individuals.

[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 3 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 2 when the total number is under 8, and divide the total number of deaths by 3 when the number is under 12. Only one property experienced a number of deaths greater than 12 (18 deaths), which we estimate affected 4 families in the household.

properties are likely to have transitioned suddenly from poor to destitute due to the death of a potential wage-earner.

Inside the catchment area, the majority of residents lived on blocks that were heavily hit. In total, 80 percent of households in the Broad Street pump catchment area lived on a block in which at least 25 percent of residences had experienced a cholera death, and 25 percent lived on a block in which at least one-half of residences had experienced a death. The corresponding figures in the rest of the parish were 10 percent and less than 1 percent. 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 percent of residents lived on a block in which at least one-quarter of residences experienced a loss, relative to only 31 percent outside of the boundary.

Based on historical accounts, the death of a wage-earner is likely to have led very quickly to changes in household behavior that produced immediate negative externalities on neighbors. Most compelling in this setting are accounts of intense crowding of both people and animals. According to historical sources, it was common for tenants trying to make ends meet to take in subtenants on short-term lease. As noted by historian Sherwell (1897, p. 35), "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, p. 28), "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 Street kept 27 dogs in a single room. He would leave (…) prodigious amounts of canine excrement to bake on the roof of the house."

It is safe to say that, particularly in such a tightly packed neighborhood, further crowding of people and animals would have led 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).

C. *Data Collection*

To track changes in real estate values, we gather several waves of data on housing prices of the roughly 1,700 housing units in St. James and construct a panel database that contains measures of property values for the years 1853 (pre-outbreak), 1864, 1894, 1936, 1995–2013, and 2015, obtained from three separate datasets. For the years 1853, 1864, 1894, and 1936, we compile data on rental prices from the National Land Tax Assessment records.[12] The Land Tax was first introduced in England in 1692 and formed the main source of government revenue until the late nineteenth century (Dowell 1965). To construct a measure of residential turnover after the epidemic, we match the names of the primary occupant at each address across the 1853 and 1864 records.

The Land Tax Assessment ended in 1963. Hence, for the years 1995–2013 we obtain property sales prices from the Land Registry of England (Land Registry 2014).[13] Records include the property address as well as the sale price and date of sale. We also obtain data on rental prices of all properties rented within the Soho area between September 2012 and May 2015 from the *LonRes* data archives, the primary source of property data for central London and only available to verified real estate agents.[14] Lastly, for the year 2015, we obtain house value estimates from Zoopla, UK's largest property listing website.[15] We digitized all valuations and addresses from the records and geocoded addresses above by matching them to housing maps from the relevant time period (for historic records) or using Google's geocoder tool (for current house records).

To assess the spatial distribution of cholera deaths, we map the total number of deaths by house using the Cholera Inquiry Committee's 1855 map (Cholera Inquiry Committee 1855), gathered immediately after the epidemic as part of a scientific investigation 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. Panel A of Figure 1 presents the resulting map, which provides information on both house locations and number of deaths per house. 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.

We also gather data on neighborhood amenities and infrastructure from maps created by scientists and city planners investigating the outbreak. 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." 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 contains the location of all 13 water pumps and public urinals, as well as neighborhood amenities including public squares, churches, police station, fire station, theaters, banks, and primary school, to which we create measures of walking distance from each residence. From

[12] Refer to online Appendix Figure SA1 for an example of the tax entries collected in 1854.
[13] Digitized data are first available in 1995.
[14] Access to the *LonRes* data archive was provided by *Greater London Properties* (LonRes 2015).
[15] "Zoopla Current Value Estimates," http://www.zoopla.co.uk/ (accessed February 20, 2015).

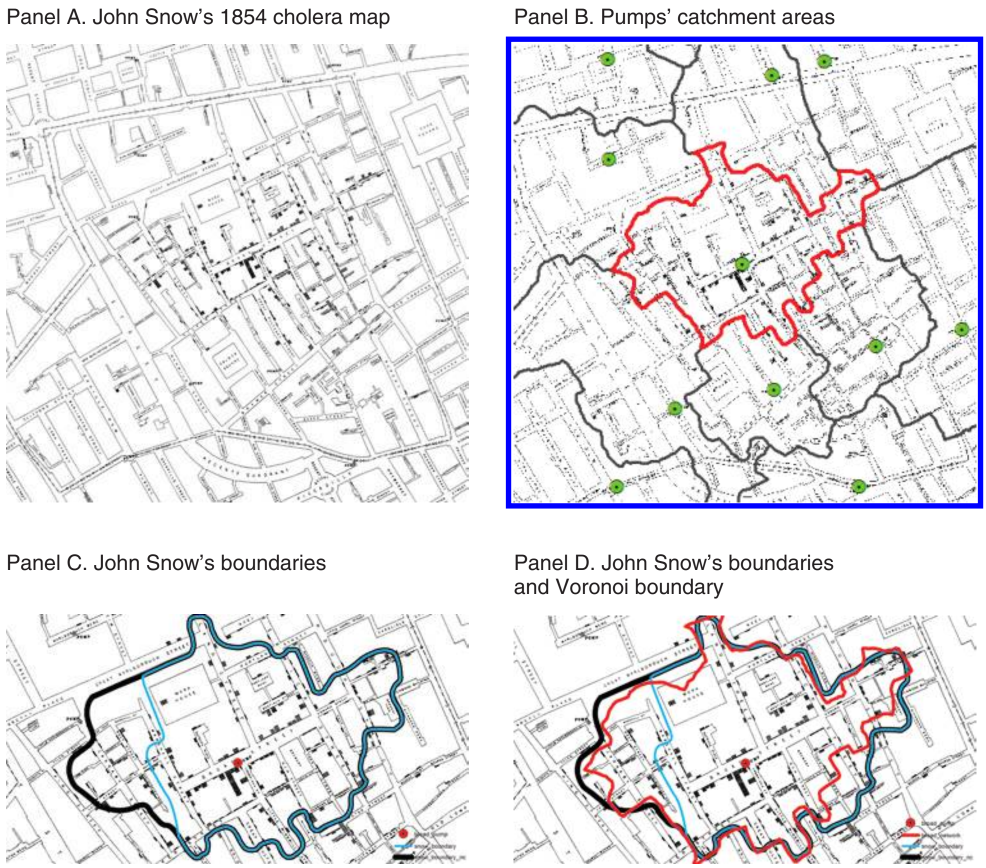


Figure 1. John Snow's 1854 Cholera Map with Pumps' Catchment Areas

Notes: Panel A: black bars denote the location of cholera deaths. Panel B: green dots indicate the location of a pump. Each catchment area is defined by a network Voronoi polygon. Broad Street pump catchment area highlighted red. Panel C: 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 Street. Panel D: adds boundary calculated by authors using Voronoi polygons (colored in red).

the same map we also calculate the distance to other neighborhood features that may influence housing prices, including the presumed and actual location of a seventeenth century plague pit (believed to be a potential source of cholera at that time), and the distance to the center of Soho. We use a similarly detailed map from 1951 to assess the location of amenities pertaining to the 1936 properties (Ordnance Survey 1951). In both cases we digitized and geocoded map data and calculated distance measures to each residence in the dataset.

We also collect data on population density and ethnic 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 microdata from Charles Booth's

1899 poverty survey of London (Booth 1902). The dataset contains household-level socioeconomic status of all households in St. James obtained from a combination of interviews with London School Board visitors and existing records collected by Board visitors during their yearly home inspections.[16] 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.[17]

II. Empirical Strategy

Property-level data allow us to assess the difference in value of properties located just inside versus just outside of the Broad Street pump catchment area using a regression discontinuity (RD) design motivated by the abrupt change in death rates at the boundary documented by Snow.

A. *Catchment Area Boundary*

As originally proposed by Snow, we define the catchment area according to a network Voronoi diagram of St. James, in which each of the 13 water pumps defines a point and the cells are determined according to the walking distance by road to each point.[18] For each property, we calculate the travel distance to each well by plotting the wells on a georeferenced 1854 street map, and then assign a property to the catchment area if the Broad Street pump is the closest well. The resulting catchment area has a total area of 0.07 square kilometers and a boundary of about 1,161 meters. A picture of the map we employ in our analysis is shown in panel B of Figure 1, alongside the original mapping of Snow in panel A. Panel B depicts the catchment areas for all pumps, and the BSP catchment area is outlined in red.[19] Cholera deaths are marked by black bars, and a portion of the map is enlarged for clarity in online Appendix Figure SA2.

As is evident from both maps, the constructed catchment area maps closely with the spatial pattern of deaths from cholera. According to our calculations, 76 percent of cholera deaths occurred within the catchment area, which is close to the

[16] 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 nonattendance, 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.

[17] 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 online Appendix Table SA1 for the original eight-tier classification used by Charles Booth.

[18] For a formal definition of network Voronoi diagrams, refer to Erwig (2000) and 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.

[19] Following previous literature (e.g., Shiode 2012) and John Snow's own accounts, we discard the pump located on Little Marlborough Street 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).

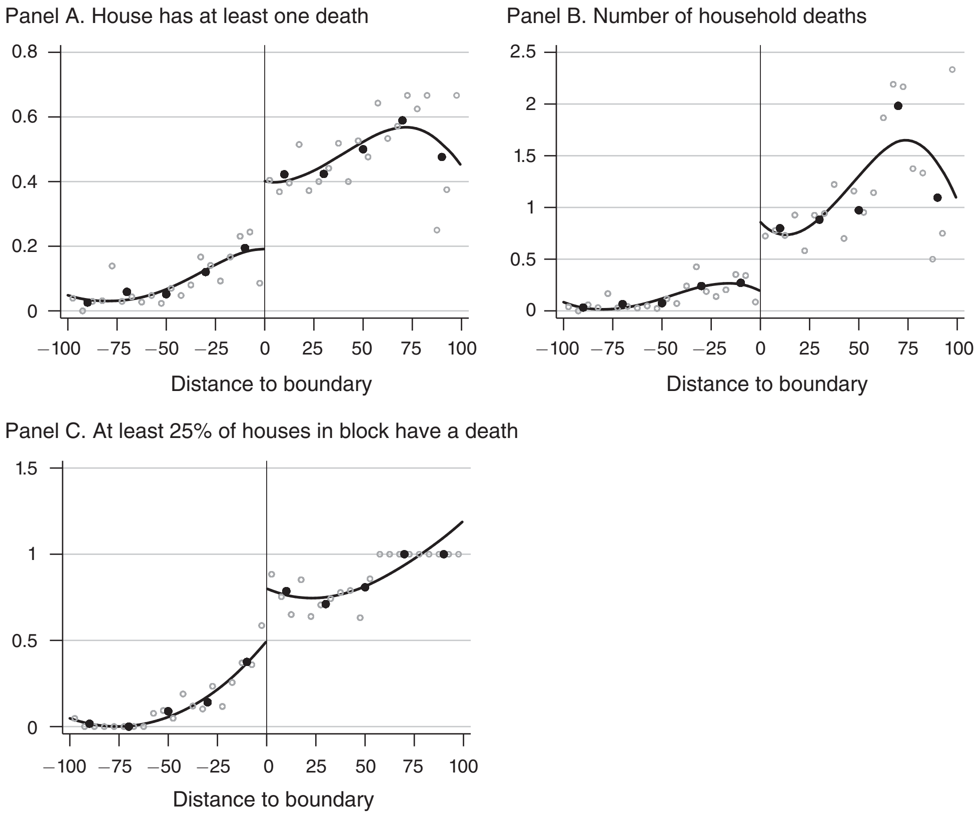


Figure 2. Cholera Deaths and BSP Boundary, 1854

Notes: Solid dots give the average value of the specified variable for houses falling within 20-meter distance bins. Hollow dots give the average value of the specified variable for houses falling within 5-meter distance bins. *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. Solid line trends are the predicted values from a regression of the specified variable on a polynomial in distance to the boundary that uses a triangular kernel and a bandwidth of 200 meters.

figures calculated by Snow. Contemporary accounts also suggest the existence of a discontinuity in deaths at the pump boundary. As Snow himself stated, "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). Panels A and B of Figure 2 plot the share of houses with at least one cholera death and the average number of deaths per household, respectively. Panel C plots the share of houses for which at least 25 percent of houses on the block experienced a death, providing further confirmation of this pattern.

This is a particularly striking pattern given that the boundaries in our setting do not determine strict assignment to particular pumps, which are nonexcludable public goods, but merely delineate the likelihood of utilizing a pump. Some residents just outside the boundary may get water from the pump because it is convenient or preferred for some reason. For instance, the Broad Street pump was located close to the local primary school, and anecdotally children drank from the pump on their way to school. Hence, while a discontinuity in deaths is not guaranteed, nor is it difficult to rationalize given existing evidence on individuals' high sensitivity to distance for water collection (Kremer et al. 2011).[20]

There is one region 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 Street to the Broad Street pump catchment area, when they are clearly closer to the Rupert Street 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 not to include them in the catchment area. Since it is possible that Snow had more accurate information on actual walking distance than we do now, in Section IIIF we verify that our estimates are robust to using the catchment area as drawn by Snow.

B. *RD Specification*

We employ a spatial regression discontinuity (RD) design that takes advantage of the discontinuity in deaths at the boundary. Given the two-dimensional nature of the Broad Street pump (BSP) boundary, we follow the usual approach in the literature by specifying a one-dimensional forcing variable, namely distance to the closest point in the BSP boundary.[21] 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.

To obtain the impact of cholera exposure on property values, we estimate the following equation restricting properties to be within bandwidth h of the BSP boundary:

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

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 the closest point on the BSP boundary; and $f(\cdot)$ is a polynomial of order K. We present results using two estimation methods for equation (1): (i) a local linear regression with RD polynomial $f(X_i) = \delta_0 X_i + \delta_1 BSP_i \times X_i$, using an optimal bandwidth h and a triangular weighting kernel.[22] (ii) A parametric specification that uses polynomial regression with RD polynomial $f(X_i) = \sum_{k=1}^{K} \delta_k X_i^k$ where the optimal choice of K is determined using Akaike's criterion as in Black, Galdo, and Smith (2007) and suggested in Lee and Lemieux (2010). For the polynomial regression, we present results both using the optimal bandwidth h from the local linear regression and using a wider, 100-meter bandwidth. 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

[20] For instance, if individuals close to the boundary use distance as a tie-breaking rule, we will observe a sharp discontinuity in deaths, particularly if habit formation leads individuals to always use the same pump.

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

[22] We follow Calonico, Cattaneo, and Titiunik (2014) for the optimal bandwidth calculation and is obtained using the rdrobust program described in Calonico, Cattaneo, and Titiunik (2015).

Table 1—House Characteristics, 1853

	Full sample			Within 100 m			RD estimates	
	Inside (1)	Outside (2)	SE (3)	Inside (4)	Outside (5)	SE (6)	RD coefficient (7)	SE (8)
Rental characteristics:								
Rental price (in logs)	3.713	3.775	(0.057)	3.709	3.740	(0.066)	0.062	(0.147)
Tax assessed (in logs)	0.446	0.515	(0.057)	0.442	0.495	(0.064)	0.044	(0.152)
Tax exonerated (yes = 1)	0.0648	0.231	(0.040)	0.0653	0.222	(0.052)	−0.034	(0.056)
Sewer access:								
Old/existing	0.472	0.565	(0.085)	0.467	0.593	(0.091)	−0.067	(0.189)
New sewer	0.401	0.275	(0.082)	0.404	0.259	(0.084)	0.248	(0.151)
No access	0.128	0.160	(0.054)	0.129	0.148	(0.061)	−0.181	(0.124)
Distance (m/100) to:								
Closest pump	1.045	0.958	(0.079)	1.051	1.065	(0.094)	0.034	(0.098)
Soho centroid	1.319	2.472	(0.118)	1.323	2.180	(0.130)	−0.152	(0.275)
Presumed plague pit	2.359	3.137	(0.224)	2.358	2.630	(0.216)	0.211	(0.401)
Public square	2.586	2.715	(0.137)	2.583	2.709	(0.141)	−0.190	(0.356)
Church	1.311	1.717	(0.130)	1.311	1.609	(0.142)	−0.184	(0.275)
Police station	4.376	5.412	(0.264)	4.379	4.792	(0.225)	0.264	(0.552)
Fire station	3.601	2.665	(0.185)	3.596	2.854	(0.224)	0.389	(0.458)
Theater	4.020	5.303	(0.236)	4.027	4.680	(0.225)	0.043	(0.512)
Pub	0.286	0.407	(0.037)	0.287	0.405	(0.046)	−0.158	(0.121)
Urinal	0.874	1.123	(0.087)	0.870	1.020	(0.088)	−0.282	(0.207)
Sewer vent	0.426	0.556	(0.047)	0.427	0.563	(0.050)	−0.166	(0.142)
Primary school	1.296	2.477	(0.132)	1.296	2.025	(0.110)	−0.368	(0.282)
Bank	3.960	4.694	(0.317)	3.958	4.086	(0.316)	0.223	(0.639)
Observations	494	1,228		490	811		559	

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 Calonico, Cattaneo, and Titiunik (2014) using a triangular kernel. Column 7 uses the average of all optimal bandwidths (29.7 meters). Number of observations for *Rental price* are 462 and 944 in columns 1 and 2, respectively; 458 and 631 in columns 4 and 5, respectively; and 513 in column 8.

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.

C. *Validity of RD Design*

Identification of the treatment effect in equation (1) requires that potential outcome functions $E[y_i(1)\,|\,X_i]$ and $E[y_i(0)\,|\,X_i]$, where $y_i(1)$ and $y_i(0)$ 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 y_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 price and neighborhood amenities. Columns 1 and 2 of Table 1 provide mean characteristics for all St. James properties inside and outside of the BSP boundary, 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, clustered at the block level.[23] Although there are significant differences in means across all properties, when we consider only properties within a 100-meter bandwidth, which is larger than most optimal bandwidths used in the empirical analysis: differences between catchment area and noncatchment area properties decrease significantly in magnitude, as expected.

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 equation (1) using a local linear regression specification and each variable in Table 1 as the dependent variable. For comparability, all estimations in column 7 use the same bandwidth (29.7 meters).[24] The patterns in Table 1 indicate that, prior to the outbreak, properties on either side of the BSP boundary are very similar: for all baseline characteristics the RD coefficients are statistically insignificant. Most importantly, measures of property value (i.e., rental prices, assessed and exonerated taxes) do not differ across the boundary even when all of St. James is considered, so significant price differences in the post-outbreak period cannot be attributed to preexisting differences.[25] Appendix Figure B1 graphically depicts the continuity of baseline covariates across the BSP boundary using averages for continuous 20-meter and 5-meter distance bins.[26]

Panel A of Table 3 verifies that rental prices in 1853 are smooth at the boundary using a local linear regression with optimal bandwidth of 35.72 meters (column 1). The results are consistent after the inclusion of a set of covariates listed in Table 1 (column 2).[27] We also verify that results are robust when using a second-degree polynomial specification of equation (1) with an optimal bandwidth (column 3) and a wider bandwidth (column 4). Lastly, we confirm that the results are consistent after including fixed effects for segments along the boundary (column 5). Specifically, we split the boundary into 5 segments of equal length (about 330 meters each) and include fixed effects for each of these segments to ensure that we are comparing house values within the same segment of the boundary.[28] Overall, the estimates in panel A of Table 3 reveal a robust absence of differences between houses inside and outside the BSP boundary with respect to real estate values prior to the epidemic. Panel A of Figure 3 graphically depicts the continuity of house values across the BSP boundary in 1853.[29]

[23] For reference, Table B1 provides Conley (1999) standard errors adjusted for spatial correlation in addition to the block-level clustered standard errors. Since Conley (1999) standard errors tend to be smaller than block-level clustered standard errors, all analyses hereafter use block-level standard errors.

[24] The bandwidth of 29.7 meters is the average of all optimal bandwidths for the variables in Table 1. Each individual bandwidth is obtained following Calonico, Cattaneo, and Titiunik (2014).

[25] Land taxes varied from year to year. However, owners could have their taxes exonerated by paying a lump sum. Once taxes were exonerated, the owners paid a yearly fixed tax that did not increase with time.

[26] Although Appendix Figure B1 depicts a visual discontinuity in sewer access, the results in this paper remain robust after accounting for this. Refer to the online Appendix for more details.

[27] Specifically, the set of covariates used are the distances to: centroid of neighborhood, public squares, fire stations, theaters, police stations, urinals, pubs, churches, banks, sewer vents, plague pit, and whether the house had sewer access.

[28] This is typically done in regression discontinuity designs with a two-dimensional boundary in order to avoid comparing units that are equally distant from the boundary but relatively far away from each other along the boundary (Keele and Titiunik 2015). In our setting, however, where the entire boundary is a little over one kilometer this is a minor concern.

[29] In an additional check, we go further back in time and combine the 1853 data with data from 1846 to assess whether there were differential pretrends in house rental values inside and outside the BSP area. As suspected, we

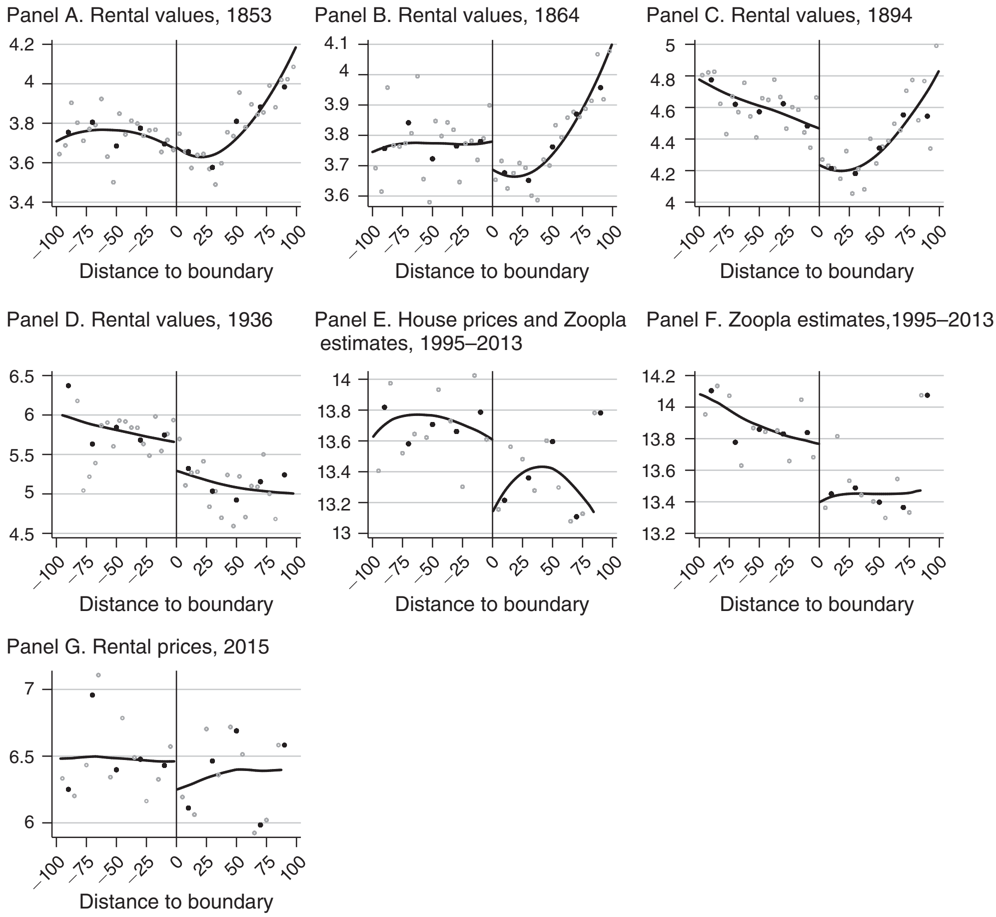


Figure 3. RD Plots for Main Outcomes (in logs)

Notes: Solid dots give the average value of the specified variable for houses falling within 20-meter distance bins. Hollow dots give the average value of the specified variable for houses falling within 5-meter distance bins. *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. Solid line trends are the predicted values from a regression of the specified variable on a polynomial in distance to the boundary that uses a triangular kernel and a bandwidth of 200 meters.

Identification of the treatment effect in equation (1) also requires that there be no endogenous sorting of properties (and individuals) within a close window around the BSP boundary. In this setting, this is a very reasonable assumption given that pumps are nonexcludable public goods, so there is no motive for individuals to sort at the boundary of the catchment area of a specific pump as there is with public school zones. The validity is also supported by quantitative assessment. In a histogram of the forcing variable (distance to the BSP boundary), there is no clear evidence of a

find no evidence of different trends in values across the two areas. Refer to the online Appendix for more details on this exercise.

Table 2—Change in Exposure to Cholera at Boundary of BSP Catchment Area

	Local linear regression		Polynomial RD		
	Baseline (1)	Controls (2)	Optimal band (3)	Wide band (4)	Segment FE (5)
Panel A. Number of deaths in household					
Inside BSP area	0.548	0.567	0.378	0.639	0.463
	(0.150)	(0.149)	(0.120)	(0.110)	(0.111)
Observations	477	485	477	1,284	1,284
Mean outside BSP	0.253	0.260	0.253	0.160	0.160
Bandwidth (meters)	25.00	25.21	25.00	100	100
Panel B. House had at least one death					
Inside BSP area	0.215	0.185	0.138	0.253	0.174
	(0.068)	(0.068)	(0.043)	(0.034)	(0.035)
Observations	538	540	538	1,284	1,284
Mean outside BSP	0.172	0.172	0.172	0.0989	0.0989
Bandwidth (meters)	28.49	28.65	28.49	100	100

Notes: Sample in *Local linear regression* specifications is restricted to be within an optimal bandwidth determined as in Calonico, Cattaneo, and Titiunik (2014) and using a triangular weighting kernel. Columns 2, 3, 4, 5 include a set of determinants of exposure to cholera as additional controls (distance to neighborhood centroid, distance to urinals, and sewer access). *Polynomial RD* specifications use a second-order polynomial in distance to the BSP boundary. *Optimal band* refers to a specification that uses the optimal bandwidth obtained from the *Local linear* specification in column 1. *Wide band* refers to a specification that uses a bandwidth of 100m around the BSP boundary (this bandwidth encompasses almost all observations within the BSP area). *Segment FE* refers to a specification that adds a set of boundary specific fixed effects that denote the closest of five boundary segments to a given observation. Clustered standard errors by street block are shown in parentheses.

jump in the density of properties across the treatment boundary (Appendix Figure B2, panel A). A McCrary (2008) test for breaks in the density of the forcing variable verifies that the density does not change discontinuously across the boundary (Appendix Figure B2, panel B).

III. Results

We start by examining how exposure to cholera influences real estate prices and residential mobility in the short run, and then investigate how valuations evolve over the next 160 years.

A. *Change in Cholera Exposure at the Boundary*

We first employ the RD framework to investigate the change in deaths from cholera at the boundary at the house level (Table 2). We look at both outcomes, the number of deaths (panel A) and whether a household experienced a death (panel B). The results in columns 1 and 2 of panel A indicate that the average number of deaths roughly triples inside the catchment area (e.g., 0.253 deaths outside BSP versus 0.801 inside). This result remains robust regardless of the specification and bandwidth used.[30] Panel B uses an indicator for whether a household has experienced at

[30] The results throughout this paper are robust to bandwidth choice. Refer to the online Appendix for further discussion on this topic.

Table 3—Boundary Effects on Rental Prices (1853, 1864, 1894, 1936)

	Local linear regression		Polynomial RD		
	Baseline	Controls	Optimal band	Wide band	Segment FE
	(1)	(2)	(3)	(4)	(5)
Panel A. log rental prices, 1853					
Inside BSP area	0.052	0.035	−0.021	−0.041	−0.079
	(0.124)	(0.078)	(0.074)	(0.073)	(0.070)
Observations	588	469	588	1,070	1,070
Mean outside BSP	47.01	46.01	47.01	48.63	48.63
Bandwidth (meters)	35.72	27.25	35.72	100	100
Panel B. log rental prices, 1864					
Inside BSP area	−0.186	−0.186	−0.113	−0.118	−0.127
	(0.118)	(0.089)	(0.068)	(0.068)	(0.068)
Observations	469	511	469	1,047	1,047
Mean outside BSP	48.48	47.74	48.48	50.24	50.24
Bandwidth (meters)	28.11	31.38	28.11	100	100
Panel C. log rental prices, 1894					
Inside BSP area	−0.255	−0.278	−0.242	−0.218	−0.217
	(0.234)	(0.147)	(0.119)	(0.076)	(0.115)
Observations	370	365	370	799	794
Mean outside BSP	119.41	119.19	119.41	121.11	121.11
Bandwidth (meters)	29.19	28.54	29.19	100	100
Panel D. log rental prices, 1936					
Inside BSP area	−0.299	−0.375	−0.325	−0.458	−0.271
	(0.311)	(0.280)	(0.147)	(0.144)	(0.150)
Observations	221	221	221	354	354
Mean outside BSP	454.49	454.49	454.49	451.43	451.43
Bandwidth (meters)	37.26	37.30	37.26	100	100

Notes: Sample in *Local linear regression* specifications is restricted to be within an optimal bandwidth determined as in Calonico, Cattaneo, and Titiunik (2014) and using a triangular weighting kernel. Columns 2, 3, 4, 5 include a set of determinants of rental values as additional controls. For panels A and B, controls are: distance to neighborhood centroid, distance to closest: square, fire station, police station, theater, urinal, pub, church, bank, pump, street vent, presumed plague pit, and sewer access. For panel C, controls are: distance to neighborhood centroid, distance to closest: square, bank, street vent, presumed plague pit. For panel D, controls are: distance to neighborhood centroid, distance to closest: square, theater, pub, church, and bank. *Polynomial RD* specifications use a second-order polynomial in distance to the BSP boundary. *Optimal band* refers to a specification that uses the optimal bandwidth obtained from the *Local linear* specification in column 1. *Wide band* refers to a specification that uses a bandwidth of 100m around the BSP boundary (this bandwidth encompasses almost all observations within the BSP area). *Segment FE* refers to a specification that adds a set of boundary specific fixed effects that denote the closest of five boundary segments to a given observation. Clustered standard errors by street block are shown in parentheses.

least one cholera death as the outcome variable. Notice also that the likelihood of a cholera death sharply increases inside the BSP catchment area. Panels A and B of Figure 2 graphically illustrate these results using RD plots.

B. *Rental Prices*

Panel B of Table 3 presents estimates of the effect of the epidemic on the evolution of property values a decade later using log rental price in 1864 as the outcome variable. We estimate a difference in rental prices at the boundary of 18 percent using an optimal bandwidth of 28 meters, and the estimates change little when we include covariates (column 2). The inclusion of covariates, however, does improve the precision of the estimates. The pattern of results is very similar when we use a second-degree polynomial specification for equation (1) (columns 3–5): properties inside the boundary experience an 11–13 percent loss in rental value in the decade after the epidemic. The results are robust when increasing the bandwidth from 28 meters (column 3) to a wider 100-meter bandwidth (column 4). The 100-meter bandwidth includes most of the properties within the BSP area and it also allows us to have results that are comparable across years in terms of the bandwidth used.[31] Furthermore, the inclusion of segment fixed effects has little impact on the magnitude and statistical significance of the effect (column 5). Refer to panel B of Figure 3 for a plot of house values in 1864 that illustrates the discontinuous drop in values within the BSP area.

The fact that some houses experienced cholera deaths outside the BSP area (and some houses did not experience cholera deaths inside the BSP area) suggests that treatment is not strictly defined by the BSP boundary. Therefore, results using a sharp RD (i.e., using whether you are inside the BSP boundary as the treatment definition) such as those in Table 3 should be interpreted as estimating the intent-to-treat effect. With this in mind, we expand our results by presenting estimates of the effect of the epidemic on property values a decade later using a fuzzy RD design. Formally, let $BSP_{fuzzy,i}$ be an indicator equal to 1 if house i experiences at least one cholera death. Panel A of Figure 2 shows the discontinuity at the BSP boundary in treatment assignment $BSP_{fuzzy,i}$. We obtain the fuzzy RD estimate by jointly estimating

$$(2) \qquad y_{it} = \alpha^f + \gamma^f BSP_{fuzzy,i} + f(X_i) + \mathbf{W}'_{it}\beta^f + \epsilon_{it} \quad \text{for } X_i < h,$$

$$(3) \qquad BSP_{fuzzy,i} = \delta + \tau BSP_i + g(X_i) + \mathbf{W}'_{it}\mu + \nu_{it} \quad \text{for } X_i < h,$$

where the estimate of γ^f is the fuzzy RD estimate. Following Imbens and Lemieux (2008), we use the same bandwidth h for outcome and treatment equations (equations (2) and (3), respectively). Further, we let polynomials $g(X_i)$ and $f(X_i)$ have the same order in both equations (Lee and Lemieux 2010). Note that since the fuzzy RD design relies on exactly knowing $BSP_{fuzzy,i}$ (i.e., whether a specific address had a cholera death in 1854), we can only estimate this model for the 1864 data when physical addresses are very similar to 1854. For the long-term analysis, it becomes difficult to pinpoint whether a specific location had a cholera death as addresses and building footprints have changed significantly over the years.

Columns 1–3 in Appendix Table B2 present the results from jointly estimating equations (2) and (3) using log rental price in 1864 as the outcome variable. Column 1 presents results using a local linear regression specification. Columns 2 and 3 use a polynomial specification with the optimal bandwidth from Column 1 and a wider 100-meter bandwidth, respectively. Houses with at least one cholera death experience between a 52 to 55 percent drop in value one decade after the

[31] Refer to online Appendix Figure SA3 for a representation of the study area with 20-, 40-, and 100-meter bandwidths around the BSP area for reference. Given gain in precision, all results hereafter include a set of baseline controls described in each table's notes.

Table 4—Boundary Effects on Residential Mobility

	Resident is different in 1864 than in 1853					log rental price, 1864	
		Polynomial RD					
	LLR (1)	Optimal band (2)	Wide band (3)	Optimal band (4)	Optimal band (5)	Optimal band (6)	Optimal band (7)
Inside BSP area	0.075	0.101	0.103	0.094	0.010	−0.160	−0.044
	(0.083)	(0.046)	(0.040)	(0.049)	(0.066)	(0.070)	(0.072)
Household death				0.072	0.053	−0.075	−0.047
				(0.044)	(0.043)	(0.049)	(0.049)
Neighborhood deaths					0.003		−0.005
					(0.002)		(0.003)
Houses in neighborhood					−0.002		0.004
					(0.001)		(0.002)
Observations	700	915	1,312	894	894	455	455
Mean outside BSP	0.594	0.517	0.515	0.517	0.517	0.594	0.594
Bandwidth (meters)	37.98	52.82	100	52.82	52.82	28.11	28.11

Notes: Clustered standard errors by street block shown in parentheses. 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. *Household death* is an indicator for whether there was at least one cholera death in the household. *Neighborhood deaths* refers to the number of deaths in the neighborhood (excluding the household being analyzed). Optimal bandwidth determined as in Calonico, Cattaneo, and Titiunik (2014) and using a triangular weighting kernel. *Optimal band* refers to a specification that uses the optimal bandwidth used in the LLR specification. *Wide band* refers to a specification that uses a bandwidth of 100m around the BSP boundary (this bandwidth encompasses almost all observations within the BSP area). All specifications include a set of baseline controls: sewer access, and distance to street vents and urinals.

outbreak (numbers in brackets present the percent effect on rental prices in 1864). The results are robust to both the specification and the bandwidth used.

In addition to the fuzzy RD results, Appendix Table B2 also presents results from an instrumental variable approach that uses whether a house is within the BSP area and the walking distance from the house to the contaminated pump as instruments for whether the house experienced at least one death (column 4) and the number of deaths in the house (column 5). Similar to the fuzzy RD results in columns 1–3, the IV results show that houses having at least one death report a 50 percent drop in rental value by 1864. Column 5 also reports that each additional cholera death translates into about a 25 percent drop in rental value 10 years later.

C. *Migration*

What effect did cholera have on migration out of the neighborhood? Tables 4 and 5 look at linear probability estimates of residential turnover as a function of location inside the Broad Street pump catchment area. We anticipate that both impoverishment from cholera and also the change in the quality of the neighborhood would have driven residents out.[32]

[32] 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 in the Commission report as being equally distributed among primary occupants and subleasing tenants (as measured by occupancy on the first floor or upper floors of the home) (Whitehead 1854).

Table 5—Migration Patterns by Cholera Exposure: BSP Catchment Area

	Resident is different in 1864 than in 1853	
	(1)	(2)
Household death	0.271	0.330
	(0.083)	(0.093)
Neighborhood deaths	0.004	
	(0.001)	
Household death × neighborhood deaths	−0.005	
	(0.002)	
Neighborhood houses with death		0.010
		(0.003)
Household death × neighborhood houses with death		−0.015
		(0.004)
Total number of houses in neighborhood	−0.001	−0.001
	(0.001)	(0.002)
Number of deaths in household	0.014	0.014
	(0.020)	(0.020)
log total sums assessed, 1853	−0.129	−0.130
	(0.077)	(0.076)
Observations	489	489
Mean outside BSP	0.504	0.504

Notes: This table shows the marginal effects of a probit estimation, with clustered standard errors by block shown in parentheses. 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. *Household death* is an indicator for whether there was at least one cholera death in the household. *Neighborhood deaths* refers to the number of deaths in the neighborhood (excluding the household being analyzed). *Neighborhood houses with death* refers to the number of houses (excluding the household being analyzed) within a neighborhood experiencing at least one cholera death. In columns 1 and 2, the mean outcome is reported among households with no deaths both within house and within neighborhood.

Overall, the neighborhood is one of relatively high turnover: 56 percent of primary tenants of St. James parish changed residence between 1853 and 1864. Furthermore, exposure to the epidemic does appear to have increased residential turnover. As shown in column 1 of Table 4, the boundary effect on mobility is 7.5 percent at the optimal bandwidth of 38 meters, but statistically insignificant. However, when we increase the bandwidth to merely 53 meters, the difference is 10 percent and strongly significant as shown in column 2.[33]

Furthermore, the exit is not just driven by mortality. As shown in column 4, the boundary effect persists (in magnitude and significance) even after accounting for deaths at the household level. Meanwhile, as shown in column 5, the increase in mobility inside 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 6 and 7). 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

[33] Appendix Figure B3 presents the RD plot for the share of houses with a different primary tenant in 1864 than in 1853.

difference at the boundary *is* explained by the number of deaths that occurred at the neighborhood level.

In Table 5 we look at the interaction between household deaths and neighborhood deaths on the propensity to move out in order 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. We therefore focus on the Broad Street pump area where the epidemic was concentrated. As shown in the second and third rows of column 1, 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. Similarly, among cholera victims, the more houses in the neighborhood that were hit by the epidemic, the more likely they are to stay (fourth and fifth rows of column 2).

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 also 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.[34] 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 one-third of households in the BSP area, at least 15 percent 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 socioeconomic shock as a result of the epidemic stayed on for part of the decade.

D. *Residential Crowding*

We use census data from just before (1851) and soon after (1861) the epidemic to more directly examine sorting of residents by socioeconomic status onto blocks inside and outside the boundary after the epidemic. Two relevant pieces of information are available in the census that are missing from the land tax data: 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 socioeconomic status in this setting.

[34] 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 percent resided at the same address after a decade.

Table 6—Boundary Effects on House Occupancy Characteristics

	Number of occupants at address		Number of immigrant families at address		Proportion of immigrant families at address	
	LLR (1)	Optimal band (2)	LLR (3)	Optimal band (4)	LLR (5)	Optimal band (6)
Panel A. Census data (1851)						
Inside BSP area	2.591	3.131	0.089	0.085	0.060	0.020
	(2.334)	(2.997)	(0.141)	(0.098)	(0.039)	(0.026)
Observations	490	490	556	556	515	515
Mean outside BSP	13.502	13.502	0.324	0.324	0.095	0.095
Bandwidth (meters)	34.81	34.81	34.07	34.07	30.94	30.94
Panel B. Census data (1861)						
Inside BSP area	4.725	4.900	0.239	0.211	0.093	0.085
	(2.425)	(3.891)	(0.189)	(0.104)	(0.052)	(0.041)
Observations	524	524	482	482	501	501
Mean outside BSP	13.62	13.62	0.391	0.391	0.126	0.126
Bandwidth (meters)	37.38	37.38	32.01	32.01	33.86	33.86

Notes: Clustered standard errors shown in parentheses. Regressions include controls for proximity to latrines and sewage status. Immigrant families are defined by a household with a head born outside of England, Wales, or Scotland. Optimal bandwidth determined as in Calonico, Cattaneo, and Titiunik (2014) and using a triangular weighting kernel. *Optimal band* refers to the polynomial specification using the optimal bandwidth from the LLR specification. Refer to the online Appendix for results using the wide 100-meter bandwidth. Census data acquired from The National Archives of the UK: Public Record Office.

Results from these estimates are presented in Table 6. Panel A shows that population density, the number of immigrant families, and the proportion of immigrant families are smooth at the BSP boundary in 1851. However, in 1861 (panel B) a sharp increase in population density emerges within the affected catchment area. Specifically, dwellings inside the boundary have 35 percent more people than those outside the boundary, or roughly five additional residents. As shown in columns 4–6 there are nearly twice as many immigrant families inside relative to outside the boundary in 1861. Furthermore, the increase in density is accompanied by an increase in the proportion of immigrant families residing in the dwelling. For both outcomes, boundary effects persist and remain significant across most specifications and bandwidth choices.[35]

E. *Long-Term Effects*

We now examine the persistence of the difference in property values that had emerged by 1864. The first set of estimates (Table 7) looks at characteristics of residents inside and outside of the catchment area boundary using data on socioeconomic status compiled by Charles Booth in 1899. By this point in time, all of Soho had piped water, so the water pump catchment areas had truly become irrelevant features of the property market. Nevertheless, we see strong evidence of discontinuities at

[35] Refer to Appendix Figure B4 for RD plots corresponding to the outcome variables discussed in this section. Furthermore, refer to the online Appendix for a replication of the results in this section restricting the sample to Irish immigrant households: which can potentially offer a better proxy for socioeconomic status.

Table 7—Boundary Effects on House Socioeconomic Status, 1899

	Very poor		Poor		Working poor		Middle class	
	LLR (1)	Optimal band (2)	LLR (3)	Optimal band (4)	LLR (5)	Optimal band (6)	LLR (7)	Optimal band (8)
Inside BSP area	0.092	0.142	0.043	−0.012	0.067	0.035	−0.236	−0.222
	(0.050)	(0.092)	(0.047)	(0.051)	(0.044)	(0.103)	(0.062)	(0.100)
Observations	523	597	581	649	571	503	483	493
Mean outside BSP	0.220	0.233	0.250	0.232	0.378	0.347	0.164	0.168
Bandwidth (meters)	25.37	30.17	29.19	34.34	28.41	24.26	22.37	23.24

Notes: Clustered standard errors shown in parentheses using street blocks as clusters. Optimal bandwidth determined as in Calonico, Cattaneo, and Titiunik (2014) and using a triangular weighting kernel. *Optimal band* refers to the polynomial specification using the optimal bandwidth from the LLR specification. Refer to the online Appendix for results using the wide 100-meter bandwidth. All specifications include controls for distances to Soho centroid, public squares, theaters, police stations, and banks.

the border in terms of household poverty level. Averaging across specifications and bandwidths, the probability of belonging to the lowest wealth class increases by more than 10 percentage points inside the boundary, and the likelihood of being middle class (the highest wealth class present in this neighborhood) falls by almost 20 percentage points.[36]

The second set of estimates (panels C and D of Table 3) considers property assessments from the land tax assessment data in 1894 and 1936. Two points about the 1936 data are worth noting. First, for reasons that are unclear, data on almost one-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. Secondly, by 1936, 45 percent of properties have been exonerated from taxation. For both reasons, the dataset is considerably smaller in this year, although importantly, attrition for both reasons is balanced inside and outside the boundary.

As shown in panels C and D of Table 3, property values continue to be significantly higher outside relative to inside the Broad Street pump catchment area in 1894 and then in 1936. Once again, the results are similar regardless of the estimation strategy used (local linear versus polynomial regression). The local linear regression results, as before, are more precise when controls are included (column 2). Furthermore, results are robust when considering a relatively wide bandwidth (column 4), and to the inclusion of boundary segment fixed effects. In fact, the point estimates in 1936, which range from 27 percent to 38 percent depending on the specification, suggest that property values had actually diverged since 1864, although the difference in magnitude across years is not statistically significant. Refer to Figure 3 for RD plots that provide a graphical representation of the discontinuity in house values in 1894 and 1936.

In Table 8, 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

[36] Appendix Figure B5 provides RD plots for the four socioeconomic status variables.

Table 8—Boundary Effects on House Prices, Zoopla House Value Estimates, and Rental Prices, 1995–2013, 2015

	House prices and Zoopla estimates		Zoopla estimates only		House rental prices	
	LLR (1)	Optimal band (2)	LLR (3)	Optimal band (4)	LLR (5)	Optimal band (6)
Inside BSP area	−0.290	−0.383	−0.324	−0.293	−0.248	−0.150
	(0.097)	(0.22)	(0.15)	(0.15)	(0.21)	(0.069)
Observations	173	163	115	104	150	150
Mean outside BSP	1,005	1,005	1,049	1,049	746.9	746.9
Bandwidth (meters)	26.60	24.15	31.11	30.75	49.10	41.15

Notes: Clustered standard errors by postal code are shown in parentheses. Optimal bandwidth determined as in Calonico, Cattaneo, and Titiunik (2014) and using a triangular weighting kernel. All specifications include a set of determinants of house values as additional controls (columns 1–2: distance to: Soho centroid, WWII dropped bomb locations, whether property is a flat, and year of sale, columns 3–4 add controls for number of bedrooms and bathrooms, columns 5–6 add controls for month rented). *Mean outside BSP* is in thousands. *Optimal band* refers to the polynomial specification using the optimal bandwidth from the LLR specification. Refer to the online Appendix for results using the wide 100-meter bandwidth.

2013, along with current rental price estimates from the UK's largest property listing website. Together, these sources provide 1,877 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 BSP catchment area of between 15 percent and 38 percent, depending on the data source. Boundary effects on rental prices are also significantly negative. Figure 3 presents RD plots for the outcomes in Table 8.

This pattern is not only striking evidence of path dependence, but also a reassuring check on the earlier data sources. That is, 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 tax records.[37]

F. *Robustness Checks*

BSP Boundary Definition.—In Appendix Table B5, we compare our results to those using the alternative definition of the BSP catchment area boundary proposed by Snow (Snow 1855). Panel C of Figure 1 depicts Snow's original boundary (in blue) and a modification of the boundary that excludes the pump at Little Marlborough Street (in black), and panel D includes the network Voronoi boundary used in the main analysis. For all years, the percentage of houses inside the network Voronoi boundary that are also inside Snow's modified boundary is close to 100 percent. Not surprisingly, the estimated RD coefficients using Snow's boundary are qualitatively similar to the ones obtained using the network Voronoi definition.

[37] Online Appendix Table SA5 replicates the results in Table 8 omitting segments of the boundary where there is a visual discontinuity in the density of houses near the boundary. Note that, overall, the results are qualitatively similar to the previous results.

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. However, changes in amenities are not anticipated to generate a discontinuity in real estate value within the space of a block, 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 estimates are sufficiently common in this setting to raise concerns about the validity of our interpretation.

To more rigorously gauge the likelihood that our treatment effects simply reflect regular fluctuations in prices across space, we examine the distribution of differences 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). 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 two million transactions.

We proceed in three steps. First, we overlay geocoded addresses on a grid covering the Greater London area with each graticule cell having a perimeter equal to Soho's perimeter. Second, within each of these cells, we randomly draw 40 pairs of subcells with length equal to the optimal bandwidth used to calculate boundary effects on contemporary real estate prices (43 meters).[38] Given that each subcell has a length of 43 meters, we choose 40 pairs in order to approximately equal the total length of the BSP boundary (about 1,700 meters).[39] Third, for each pair, we randomly assign one of the two subcells into treatment and calculate the average difference in log prices between treated and nontreated subcells for each Soho-sized cell. Finally, 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.

Figure 4 presents the distribution of log price differences for the randomly drawn pairs of subcells. Dashed lines indicate the fifth and tenth percentiles of the distribution. The solid vertical lines give the estimated boundary effect observed in Soho and previously discussed in Section IIIE and presented in column 1 of Table 8. Note from panel A that price differences are clustered around zero indicating that, within cells, prices differ very little for paired subcells. This is not the case in our setting. In fact, the close to 30 percent drop in prices at the BSP boundary observed in Soho is well below the fifth percentile of the average price differences observed in paired subcells across London. The evidence remains robust if the distribution is limited to cells with average prices within the range observed in Soho (panel B).

[38] Refer to Table 8 for the bandwidths used to calculate boundary effects on contemporary real estate prices.

[39] For a given subcell, a pair is defined as that subcell and the subcell that is directly adjacent to it. In cases where a subcell had four adjacent subcells, we randomly draw one of the four to serve as the pair for that cell. Note that only subcells with more than one house within its area are included in the analysis.

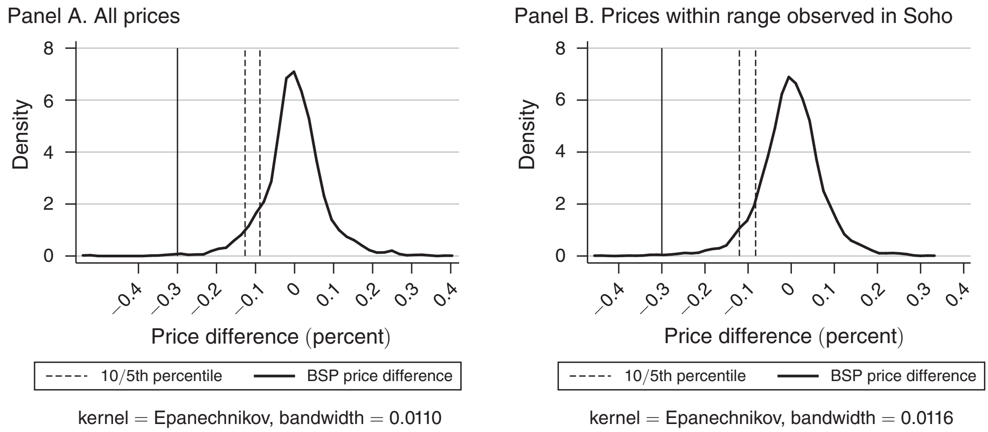


Figure 4. Difference in log House Prices, Neighboring Cells, London, 1995–2013

Notes: Histogram of log price differences for houses in adjacent subcells for the Greater London area. Dashed lines give the fifth and tenth percentile of the distribution. Solid lines give the estimated price difference between treated and nontreated houses across the BSP boundary in Soho. Each cell determined using a grid covering the Greater London area with each graticule cell having length equal to the perimeter of Soho. Each cell is divided into subcells with length equal to the bandwidth used in the main house price analysis in the Soho area (43 meters). Sample excludes cells within Soho area and cells intersected by a primary road, railway, monorail, river, or canal and cells with a single observation. Grid determined using "Create Fishnet" tool in ArcGIS.

In a related exercise, we replicate the RD design using the catchment areas of pumps that were not the source of the cholera outbreak. Figure 5 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 5 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.

Panels A through D in Appendix Table B3 present the RD coefficients from the estimation of equation (1) using each of the four boundaries in Figure 5 as the treatment boundaries. We use false boundaries 1, 2, and 3 in the pre-outbreak (1853) and post-outbreak (1864, 1894) periods (columns 1–6), and false boundaries 2 and 4 for the exercise using current property values (columns 6 and 7).[40] Similar to the pre-outbreak results at the BSP boundary, there is no evidence of a pre-outbreak "pump effect" for any of the tested pump catchment areas. 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 verify that no significant changes in deaths occur at the boundaries of unaffected pumps.[41] Columns 4–8 confirm that property prices in 1864 and 1894, migration from 1853 to 1864, and current house values are not significantly different across the false boundaries. We note that the coefficient for 1894 rental prices in Boundary 2 happens to be significant, however, this is likely due to sampling error given the consistency in the results for the other boundaries tested.

[40] Because of data availability, we cannot perform any falsification tests for the 1936 data.
[41] The coefficient on boundary 3 for whether a house has at least one death is slightly significant, however, this is likely due to sampling error given the consistency in the results for the other boundaries and the other cholera measure (column 2).

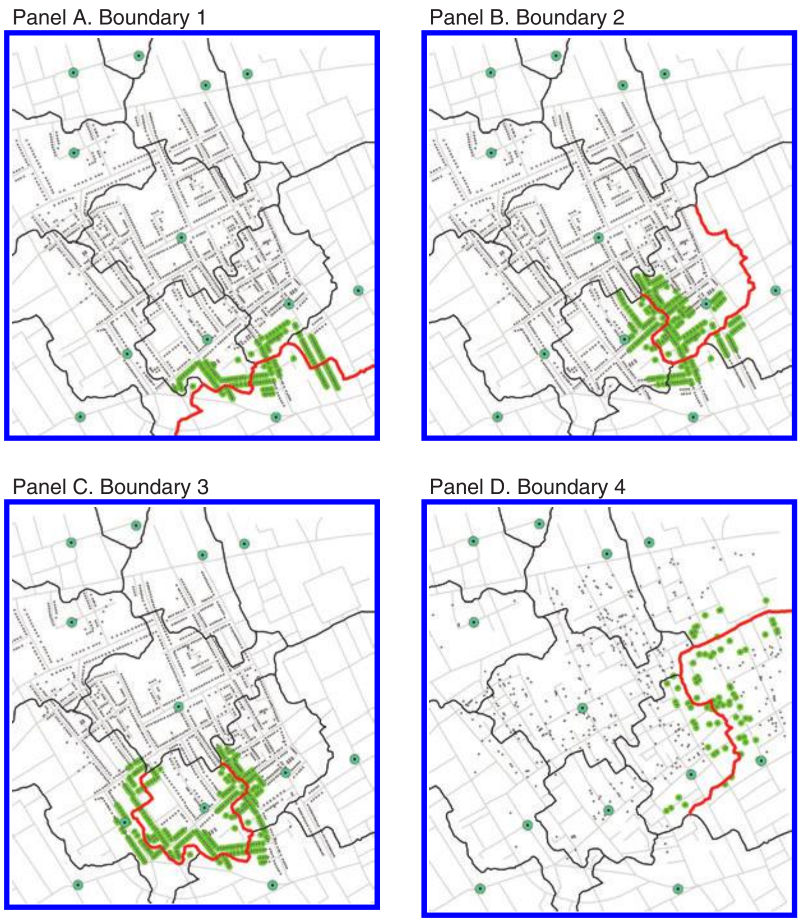


Figure 5. False Treatment Boundaries and Estimation Samples

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 Calonico, Cattaneo, and Titiunik (2014).

G. *Possible Interpretations*

Before detailing the proposed mechanism underlying changes in real estate prices at the border that is outlined in the introduction, we discuss two alternative interpretations for the patterns that emerged.

Stigma of Disease.—Is it 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 deaths from cholera carried a stigma that reduced renters' willingness to live inside the catchment area? While it is impossible to rule out any role of stigma, two pieces of evidence suggest that stigma is not entirely responsible for the differences observed. 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 (panel B of Table B4). Hence, there is no evidence of property-specific stigma effects.[42]

Second, the RD design 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 would be difficult for media sources to reference specific blocks, making it less likely that stigma changes sharply along an invisible boundary within the space of a couple of blocks.

It is also worth noting that 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, can be absorbed into our model as simply another source of disutility prospective renters experience from living among the poor, and hence do not change our long-run predictions on boundary effects.

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 12 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 160 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 nineteenth century, piped water had replaced well water in Soho, so property prices could not possibly reflect updated beliefs about well quality.

IV. Theoretical Framework

The empirical findings suggest a "tipping" story, according to which if a shock to a block is severe enough, it pushes it over a threshold, and the block tips into an equilibrium with relatively poorer tenants. On the other hand, if the shock is moderate, the block converges back to the original equilibrium. Our setting provides some additional challenges for explanations along the lines above. The first one is the nature of the shock, namely that it only directly affected people, not infrastructure. Second, the small size of the affected area and that affected blocks are in close proximity and relatively intertwined with unaffected blocks. Third, apartments in the affected blocks were in the hands of landlords who by screening tenants via prices could potentially control the composition of blocks. These features mitigate various

[42] Appendix Table B4 also presents results using the number of deaths in the house instead of whether a house has experienced a cholera death as the main explanatory variable. However, both measures yield similar results.

standard channels that typically lead negative shocks to persist over time. For example, both types of blocks in our analysis belong to the same local jurisdiction, school districts, etc., so the explanation of different social institutions evolving, cementing initial differences, is not a possibility in our setting.

In the Appendix, we provide a model demonstrating that tipping can occur even in a setting with the features above. The key elements of the model are (i) a block-level negative externality from living near poor neighbors, (ii) a higher willingness to pay to avoid this externality for the rich than the poor, and (iii) some friction that imposes a cost on landlords when they change tenants in a block. With these assumptions, we show that for a range of the parameter space (in particular when the frictions associated with changing tenants are not too small and the landlords' discount factor is not too high), although the landlords could make pricing decisions that would take back the block to its original tenant composition, if the initial shock is too severe, it would be too costly for them, and instead it is optimal for them to let the block transition to having poorer tenants.

The model assumes that the main direct effect of the cholera epidemic was increasing the share of poor tenants in affected blocks through the death of wage-earners. Other direct effects, such as the epidemic temporarily reducing local amenities in affected blocks, or inducing rich tenants having higher willingness to pay for such amenities to leave, are possible. However, we do not focus on these because of the close proximity of affected and unaffected blocks (suggesting that differences with respect to access to local amenities were small).[43] As for element (iii), the model assumes that it comes from contracting frictions (landlords have to wait until existing tenants' contracts expire before kicking them out) and that in the meantime they have to offer discounts for rich tenants to be willing to stay in a block with poor neighbors. Other types of frictions are also possible, as well as coordination issues among multiple landlords within a block. We show in the model that even without the coordination problem (that is, with a single landlord within a block) tipping is possible. Furthermore, in an extension of the model, we show that with multiple landlords, and hence likely coordination issues, this scenario is more likely.

In extensions of the base model we show that the following factors make the tipping scenario more likely: (i) smaller investment/maintenance costs associated with housing with poor versus rich tenants, (ii) multiple landlords within a block and associated coordination issues, (iii) if landlords have more restricted ability to price discriminate than in the base model, and (iv) relative income differences can stay persistent even when the whole geographic area (both affected and unaffected blocks) gentrifies and becomes richer, as in the case of Soho in the past few decades. Lastly, we provide back-of-the-envelope calculations showing that with realistic parameter values the model implies that the threshold for tipping is when about 40 percent of the households in a block are affected by the epidemic, which is consistent with our data.

We note that the model investigates transition within a block from rich to poor in isolation. The theoretical exercise is thus valid as long as cross-block externalities

[43] We do consider an extension of the model though in which over time differences evolve between buildings in affected versus unaffected blocks, as owners might invest differentially in maintaining buildings with poor versus rich tenants.

(tenants caring about the composition of tenants in neighboring blocks, as opposed to in own block) are relatively small. Large cross-block externalities would necessitate a much more complicated general equilibrium analysis, comprising all blocks.

A. *Related Literature and Contribution*

There are various papers in the literature investigating transitions of neighborhoods from high income to low income that are closely related to our model. In a classic paper, Schelling (1969) makes the point that if the integration of a neighborhood reaches a critical level, the neighborhood can tip and suddenly change from relatively high income to low income.[44] Schelling's model is stylized, as it involves no prices, and the agents in the model are not fully forward-looking. Anas (1980) provides a theoretical analysis of changes in a neighborhood's social mix to gradual exogenous changes that includes a housing market with prices, and characterizes conditions under which a sudden tipping can occur. The main differences between Anas (1980) framework and ours are that (i) the former focuses on exogenous changes in housing quality, and structural and amenity characteristics of a neighborhood, while we focus on an exogenous change in the social composition; and (ii) Anas (1980) assumes a perfectly competitive housing market, while we postulate price-setting landlords, who have the means to revert the composition of renters to the original state, but it is costly for them to do so because of inertia and the discounts they need to offer to rich tenants in the transition period.[45]

Lee and Lin (2018) also features a similar model, but it focuses on how geographic features can anchor neighborhoods, making transitions with respect to social mixture less likely.[46] Furthermore, while Lee and Lin (2018) assumes that individuals make a separate neighborhood choice every period (perfect mobility), inertia and forward-looking agents play a central role in our model.[47] To summarize, our model demonstrates that neighborhood tipping dynamics similar to those described in the existing literature are possible even when there is no initial change in physical amenities, and when prices are controlled by price-setting and fully forward-looking landlords (in the extreme, a monopolist landlord).

Our work is also related to Hornbeck and Keniston (2017), which analyzes the impact of the Great Boston Fire of 1872 on housing prices. Their theoretical framework considers a dynamic model of urban development. They show that in the presence of neighborhood externalities, affected neighborhoods can gain economic benefits afterward through the destruction forcing synchronized development in the area, and this is indeed what they find empirically. The main difference between the two settings is that, as opposed to the Boston fire, the cholera epidemic in London

[44] See also Schelling (1971, 1978); Brock and Durlauf (2001); Glaeser and Scheinkman (2000); Becker and Murphy (2009).

[45] Card, Mas, and Rothstein (2008) also features a simple competitive model framework, and they empirically verify tipping behavior in the dynamics of neighborhood racial composition.

[46] Since the area we examine is homogeneous with respect to natural amenities, the finding that blocks strongly affected by cholera transitioned to poor is consistent with the predictions of Lee and Lin (2018). However, their model does not explain why relative differences between the affected and unaffected blocks remained stable for a very long time. Our model addresses this.

[47] For an earlier, static modeling approach for household location choice within a city, see Brueckner, Thisse, and Zenou (1999).

did not impose direct infrastructure damage, hence the mechanisms that Hornbeck and Keniston (2017) focuses on are different in our setting. The cholera epidemic directly affected the income potential of many families in affected blocks, and subsequently it also could have induced less investment or less maintenance by landlords in these buildings (which we incorporate as an extension into our model). But it did not induce the type of synchronized reconstruction as the Boston fire did, which is perhaps why we find the opposite impact on housing prices than Hornbeck and Keniston (2017).[48]

More similar to our investigation is Heblich, Trew, and Zylberberg (2016), which shows the persistence of the effect of historic pollution on housing prices in English cities after the second Clear Air Act of 1968. They study a shorter time period than ours (1971–2011), but the main difference is that their unit of analysis (Lower Super Output Areas (LSOAs)) are larger than blocks, our unit of analysis. In particular in the context of Heblich, Trew, and Zylberberg (2016), there is more scope for different amenities (local institutions, local infrastructure, etc.) evolving in affected versus nonaffected LSOAs during the long period of air pollution. This introduces additional channels through which the persistence of the housing price differences can be explained. In our setting, there is much less scope for this, as affected and unaffected blocks belong to the same school district, administrative district, etc.

More generally, the idea that individuals self-sort via mobility to different communities goes back to Tiebout (1956), a paper that inspired a large literature on interjurisdictional mobility.[49] Our paper differs from this literature by focusing on mobility in different blocks that belong to the same jurisdiction. There is also a large literature on changes in the internal structure of cities, investigating phenomena like the decentralization of US cities and later the gentrification of some city centers in the twentieth century.[50] Our investigation focuses on changes within an even smaller geographic area.

Our model has the feature that the same apartment is rented cheaper to a high-income than to a low-income renter, the difference being a premium to attract the rich renter to a currently mixed composition block. This is an old hypothesis in the related literature, and it received empirical verification in several works.[51] Landlords in our model choose this strategy because we assume that individuals value having richer neighbors (hence attracting rich residents increases the rents the landlord can charge in the future). Recent theoretical papers similarly assuming that neighborhood composition directly enters individuals' utility functions include Guerrieri, Hartley, and Hurst (2013) and Lee and Lin (2018). Bayer, Ferreira, and McMillan (2004, 2007) and Bayer, McMillan, and Rueben (2004) model preferences for other residents either directly or indirectly via public services. For a more general model framework for social interactions on economic outcomes, see Manski (1993, 2000) and related discussions in Ioannides (2013).

[48] See also Siodla (2015) and Dericks and Koster (2016) for papers examining the effect of historic events causing direct infrastructure damage to some parts of a city, with similar findings as in Hornbeck and Keniston (2017).

[49] Some of the influential papers in this literature are Oates (1969); Bergstrom and Goodman (1973); Bergstrom, Rubinfeld, and Shapiro (1982); and Epple and Sieg (1999).

[50] See, for example, Jackson (1987), Mieszkowski and Mills (1993), or Aaronson (2001).

[51] For example, Muth (1969), Quigley (1974), and Kain and Quigley (1975).

V. 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, demographic trends could play a role similar to investment 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 into slightly lower-priced blocks, the emergence of enclaves 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 licensing procedure for sexual entertainment venues (SEVs), many of which have historically been located in Soho. The city council has full jurisdiction over which establishments are granted licenses, which could lead to the sorting of SEVs onto relatively impoverished neighborhood blocks. Assuming such establishments generate negative externalities on residents of the same block, the spatial sorting of SEVs could contribute to persistent differences in income levels across neighborhood blocks. The transaction cost of the licensing procedure further suppresses block-level gentrification by making it especially costly for firms to relocate.

Related are zoning codes that were introduced in the twentieth century, which regulate how specific properties are used and inhabited. Although the United Kingdom did not introduce a formal system of zoning as did the United States, various Town Planning Acts introduced urban planning legislation beginning in 1909. Most relevant, the 1947 Town and Country Planning Act required local authorities to develop Local Plan Maps outlining what kind of development out of 14 distinct use classes was permitted on each lot. Property use designations can vary within a small area, and, like commercial licenses, are costly to reassign. If blocks impoverished by the cholera epidemic, or properties within those blocks, were in the twentieth century systematically designated with development plans allowing for more commercial use, for example, then border effects would be more likely to persist over time. Relatedly, Brooks and Lutz (2019) documents a similar source of persistence in housing prices in the United States due to zoning patterns that are endogenous to property values.

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. 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. Nonetheless, even excluding this period entirely, our results provide evidence of path dependence in property values over 90 years in a relatively unregulated property market.

VI. 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 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.

Appendix A. Theoretical Analysis

We show that a simple model of sorting and neighborhood externalities can explain the persistence we observe. In particular, we build a model of a rental market in which an unexpected and localized negative income shock hits some tenants on particular blocks, and characterize conditions under which such a shock can permanently change the composition of renters.

A. *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. We consider a discrete time model with time periods $t = 0, 1, 2, \ldots$, 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. Let δ denote the common discount factor between periods. We assume that, after the shock, $x \in \{0, \ldots, n-1\}$ 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 feature of this setup is that the composition of the block can only change gradually. 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. 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 $k \in \{0, \ldots, n-1\}$: that is, poor tenants' willingness to pay to reduce the number of poor neighbors is zero. But in the online Appendix, we show 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. 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 factor δ. In the baseline model we allow the landlord to perfectly screen potential new tenants and essentially choose the type of tenant. 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 nonexogenous 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.

B. *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 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 online 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 if and only if*

$$W^r - W^p > \sum_{i=0}^{x} \frac{(1-\delta)\frac{x!}{i!}(x+1-i)\left(\delta\frac{q}{n}\right)^{x-i}\left(1-\delta\left(1-\frac{q}{n}\right)\right)}{\prod_{m=i}^{x+1}\left(1-\delta\left(1-q\frac{m}{n}\right)\right)} c_i^r. \tag{A1}$$

Note that an increase in W^r relative to W^p increases the left-hand side of (A1), and so increases the threshold x^* at which the landlord switches to the always poor strategy. The right-hand side of (A1) is increasing in each c_i^r $(i \in \{0,\ldots,x\})$, 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 (A1) 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. 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.

For the derivation of condition (A1), see the online Appendix. As some of the extensions we consider below, for tractability, assume $n = 2$, here we provide the formula for the case of $n = 2$, when the existing tenant at the start of the game is a poor type:

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

C. *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. Assume that a cost-to-disutility ratio k/d is low enough that in case of the "always rich" strategy it is profitable to always choose H. In the online Appendix we show that this is equivalent to assuming that

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

If this condition holds then in the case when the 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 following an always rich strategy decreasing to

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

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

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, but this requires a more demanding condition than the one implying that the all rich strategy is optimal for a single 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 profile above 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 strategy profile above, the landlord's expected payoff is

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

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

$$\text{(A2)} \qquad W^r - W^p > \frac{\delta\frac{q}{2}}{1-\delta\left(1-\frac{q}{2}\right)}c_0^r + \frac{1-\delta}{1-\delta\left(1-\frac{q}{2}\right)}c_1^r.$$

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 (A2) holds, there might be another equilibrium, caused by the coordination problem between the two landlords, in which both landlords follow the all poor strategy. In the online Appendix, we show that such equilibrium can be ruled out if and only if

$$\text{(A3)} \qquad W^r - W^p > \frac{1-\delta+\frac{1}{2}\delta q}{1-\delta+\delta q}c_0^r + \frac{\frac{1}{2}\delta q}{1-\delta+\delta q}c_1^r.$$

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 online Appendix we show that in this modified environment the always rich strategy yields a higher payoff for the landlord than the always poor strategy if

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

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 toward wealthier tenants. Such a trend characterizes Soho over the last two decades, during which time average sales prices have increased by 139 percent. Meanwhile, our empirical results indicate that the wedge in rental prices remains even as the district has gentrified.

In the online Appendix, we extend our model to include four types of renters: poor, middle-class, rich, and very rich, and 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 upward in both of the blocks, due to an exogenous increase in the attractiveness of the blocks.

D. *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 will choose 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), we set the returns to capital around the time of the cholera epidemic to 20 percent, hence the relevant discount factor for a landlord to $1/1.2$. 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 full block transitions would have finished with very high probability) is 26 percent, 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 percent). We set the time between periods to be a week, and set q such that contracts on average get renegotiated every 2 years.

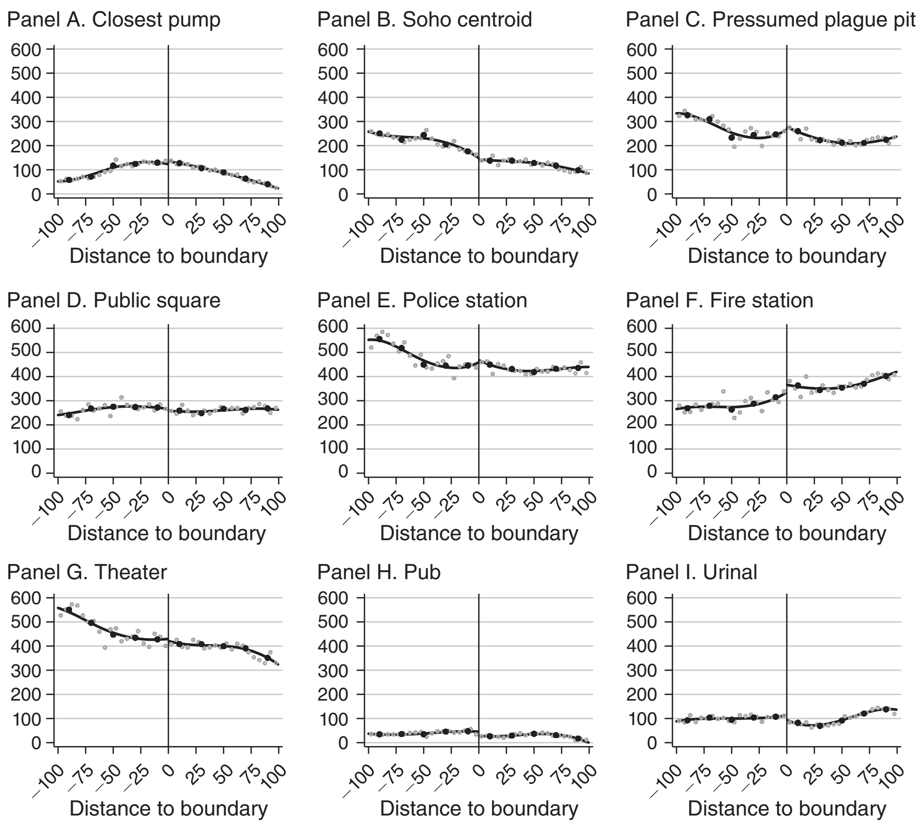
Assuming a 6 percent interest rate we find that the minimum incremental disutility rationalizing the owner choosing the all poor strategy when 40 percent of tenants are hit by the shock is $y = 0.102$, when the all rich strategy would require offering an initial rent of $r = 0.23$ (77 percent discount relative to a poor tenant's rent) to the very first new rich person moving in. Assuming a 5 percent interest rate increases the minimum incremental disutility to $y = 0.121$. Considering a block in which 50 percent of households are hit, the minimum incremental disutility changes to 0.081 with 6 percent interest rate, and to 0.097 with 5 percent interest rate.

The levels of disutilities from poor neighbors, and relatedly, the amount of discount a landlord has to offer to retain rich tenants, are not unreasonable values. 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 aspects above decrease the threshold incremental disutility parameter (and also the initial discount required to retain a rich tenant). We also did computations with a strictly concave disutility function (the very first poor neighbor imposes 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 computations above are based on a relatively low discount factor of $1/1.2$. Over time, the returns to capital decreased to levels significantly lower than the 20 percent 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 parameter values of the model above, 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 percent. Moreover, if we assume that the high return to capital prevailed throughout the period of heavy industrialization, until around 1915, then even when assuming that landlords in 1860 foresaw the drop of capital returns after this period to 5–10 percent, 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 interest rates above it only has a minuscule impact on the landlords' expected payoffs.

These calculations lend credibility to our interpretation by verifying that blocks transitioning to indefinite poverty once enough residents become poor is not only a theoretical possibility, but one that can occur with reasonable parameter values.

Appendix B. Additional Tables and Figures



(*Continued*)

Figure B1. Covariate RD Plots (1853)

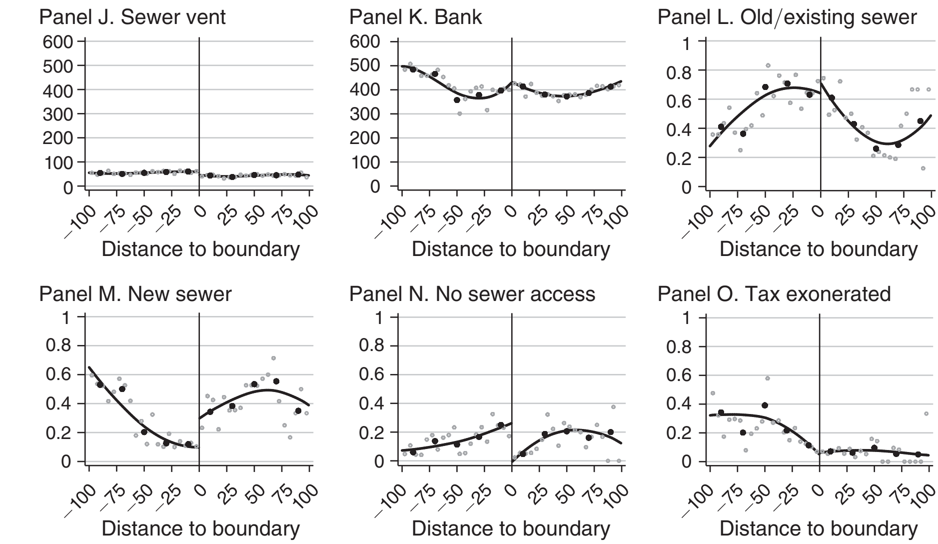


Figure B1. Covariate RD Plots (1853) (continued)

Notes: Solid dots give the average value of the specified variable for houses falling within 20-meter distance bins. Hollow dots give the average value of the specified variable for houses falling within 5-meter distance bins. *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 polynomial in distance to the boundary that uses a triangular kernel and a bandwidth of 200 meters.

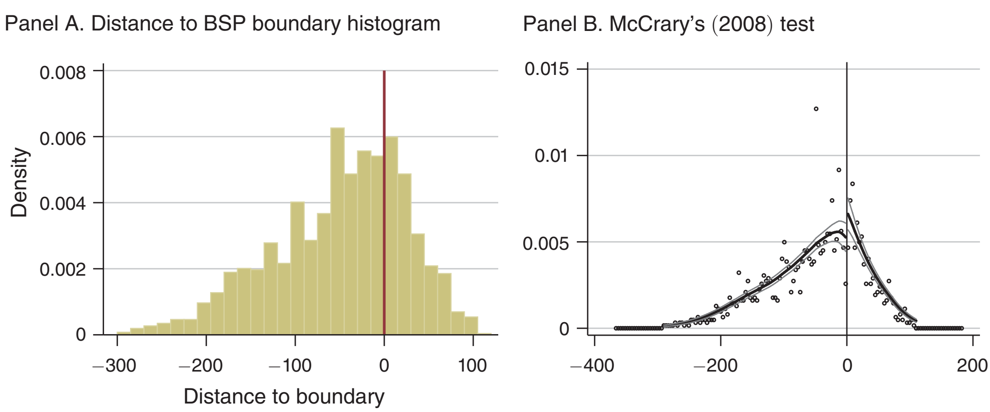


Figure B2. Density of Forcing Variable (Distance to BSP Boundary)

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.

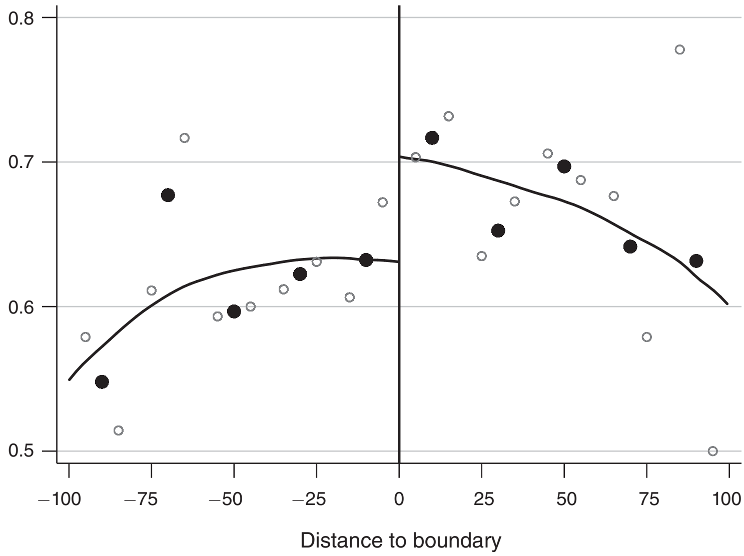


Figure B3. RD Plots for Residential Mobility Outcome

Notes: Vertical axis denotes the share of houses where head of household in 1864 is different than in 1853. Solid dots give the average value of the specified variable for houses falling within 20-meter distance bins. Hollow dots give the average value of the specified variable for houses falling within 10-meter distance bins. *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. 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 triangular kernel and a bandwidth of 200 meters.

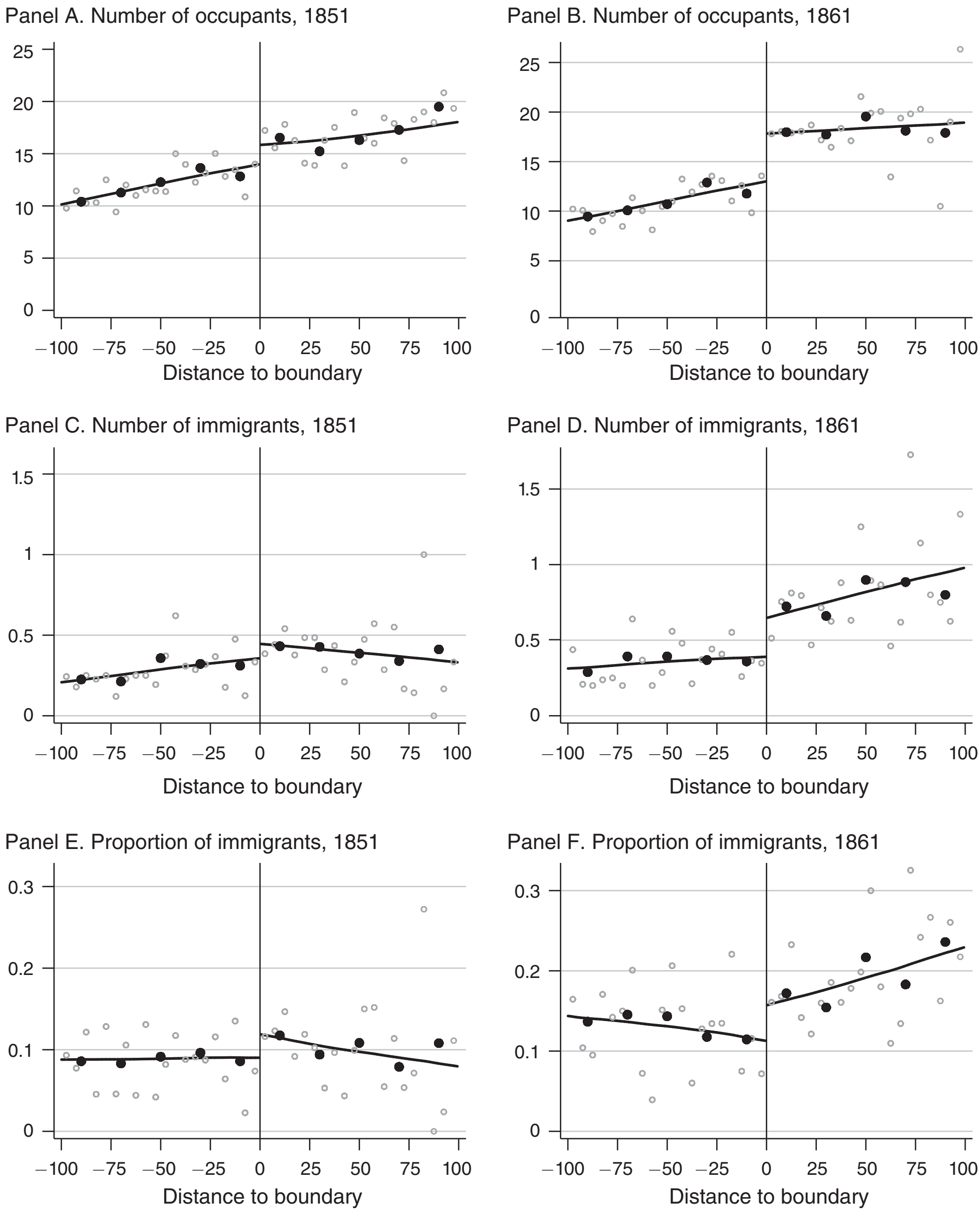


Figure B4. RD Plots for House Occupancy Outcomes

Notes: Solid dots give the average value of the specified variable for houses falling within 20-meter distance bins. Hollow dots give the average value of the specified variable for houses falling within 5-meter distance bins. *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. Solid line trends are the predicted values from a regression of the specified variable on a polynomial in distance to the boundary that uses a triangular kernel and a bandwidth of 200 meters.

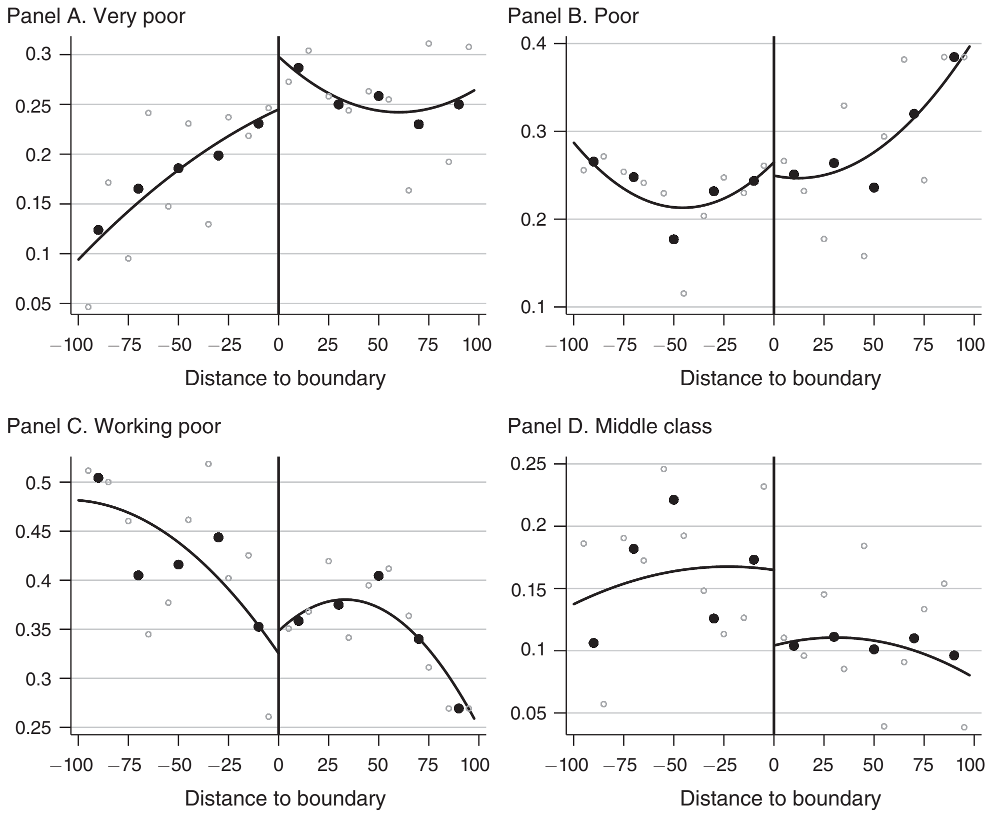


Figure B5. RD Plots for Socioeconomic Outcomes

Notes: Solid dots give the average value of the specified variable for houses falling within 20-meter distance bins. Hollow dots give the average value of the specified variable for houses falling within 10-meter distance bins. *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. Solid line trends are the predicted values from a regression of the specified variable on a polynomial in distance to the boundary that uses a triangular kernel and a bandwidth of 200 meters.

Table B1—House Characteristics (1853) with Conley (1999) Standard Errors

	Full sample			Within 100 m		
	Inside (1)	Outside (2)	SE (3)	Inside (4)	Outside (5)	SE (6)
Rental characteristics:						
Rental price (in logs)	3.713	3.775	(0.057) [0.057]	3.709	3.740	(0.066) [0.064]
Tax assessed (in logs)	0.446	0.515	(0.057) [0.056]	0.442	0.495	(0.064) [0.063]
Tax exonerated (yes = 1)	0.065	0.231	(0.040) [0.039]	0.065	0.222	(0.052) [0.051]
Sewer access:						
Old/existing	0.472	0.565	(0.085) [0.086]	0.467	0.593	(0.091) [0.092]
New sewer	0.401	0.275	(0.082) [0.084]	0.404	0.259	(0.084) [0.087]
No access	0.128	0.160	(0.054) [0.053]	0.129	0.148	(0.061) [0.058]
Distance (m/100) to:						
Closest pump	1.045	0.958	(0.079) [0.078]	1.051	1.065	(0.094) [0.092]
Soho centroid	1.319	2.472	(0.118) [0.115]	1.323	2.180	(0.130) [0.128]
Presumed plague pit	2.359	3.137	(0.224) [0.213]	2.358	2.630	(0.216) [0.212]
Public square	2.586	2.715	(0.137) [0.134]	2.583	2.709	(0.141) [0.143]
Church	1.311	1.717	(0.130) [0.129]	1.311	1.609	(0.142) [0.143]
Police station	4.376	5.412	(0.264) [0.252]	4.379	4.792	(0.225) [0.226]
Fire station	3.601	2.665	(0.185) [0.184]	3.596	2.854	(0.224) [0.221]
Theater	4.020	5.303	(0.236) [0.228]	4.027	4.680	(0.225) [0.226]
Pub	0.286	0.407	(0.037) [0.036]	0.287	0.405	(0.046) [0.044]
Urinal	0.874	1.123	(0.087) [0.087]	0.870	1.020	(0.088) [0.089]
Sewer ventilator	0.426	0.556	(0.047) [0.048]	0.427	0.563	(0.050) [0.051]
Primary school	1.296	2.477	(0.132) [0.127]	1.296	2.025	(0.110) [0.113]
Bank	3.960	4.694	(0.317) [0.301]	3.958	4.086	(0.316) [0.311]
Observations	494	1,228		490	811	

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 parentheses 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 uniform spatial weighting kernel (as recommended in Conley 2008). The choice of distance for the kernel is 53 meters, the average length of a block in the sample. *Inside* and *Outside* indicate whether a house is inside or outside the BSP area, respectively. Number of observations for *Rental price* are 462 and 944 in columns 1 and 2, respectively; and 458 and 631 in columns 4 and 5, respectively.

Table B2—Fuzzy RD and IV Estimates of Boundary Effects on Rental Prices, 1864

	Fuzzy RD estimates				
	LLR	Polynomial RD		IV estimates	
	(1)	(2)	(3)	(4)	(5)
At least one death	−0.799	−0.753	−0.767	−0.694	
	[−0.550]	[−0.529]	[−0.536]	[−0.501]	
	(0.483)	(0.392)	(0.719)	(0.326)	
Number of deaths					−0.293
					[−0.254]
					(0.141)
Observations	512	512	1,024	467	467
Mean outside BSP	48.47	48.47	50.24	48.23	48.23
Bandwidth (meters)	33.02	33.02	100	29.20	29.20
First-stage F-statistic	—	—	—	10.58	11.34
Sargan-Hansen p-value	—	—	—	0.156	0.161

Notes: LLR refers to Local Linear Regression. Sample in *LLR* specifications is restricted to be within an optimal bandwidth determined as in Calonico, Cattaneo, and Titiunik (2014) and using a triangular weighting kernel. All specifications include a set of determinants of house values as additional controls (distance to: Soho centroid, closest urinal, closest pump, and sewer status). *Polynomial RD* specifications use a polynomial in distance to the BSP boundary. Order of polynomial determined using Akaike's criteria as suggested in Black, Galdo, and Smith (2007). Column 2 uses the optimal band used in the LLR specification. Column 3 uses a wide band of 100m around the BSP boundary (this bandwidth encompasses almost all observations within the BSP area). Numbers in brackets transform the estimated coefficient $\hat{\beta}$ using the expression $\exp(\hat{\beta}) - 1$. Clustered standard errors by street block are shown in parentheses.

Table B3—False Treatment Boundary Tests

	Pre-outbreak	Cholera exposure		Post-outbreak			Current, 1995–2013, 2015	
	Rental price, 1853 (1)	Number of deaths (2)	House has at least one death (3)	Rental price, 1864 (4)	Different resident in 1864 (5)	Rental price, 1894 (6)	Price and Zoopla estimates (7)	Zoopla estimates (8)
Panel A. False boundary 1								
RD coefficient	−0.015	−0.051	−0.010	−0.030	0.325	−0.204		
	(0.120)	(0.046)	(0.064)	(0.166)	(0.188)	(0.326)		
Observations	96	114	114	130	155	68		
Bandwidth (meters)	26	25	25	37	35	53		
Panel B. False boundary 2								
RD coefficient	−0.221	−0.031	0.013	−0.290	−0.201	−0.428	−0.068	−0.010
	(0.145)	(0.101)	(0.079)	(0.217)	(0.214)	(0.173)	(0.127)	(0.086)
Observations	217	144	179	136	148	109	285	95
Bandwidth (meters)	46	27	34	30	28	42	42	32
Panel C. False boundary 3								
RD coefficient	−0.098	−0.041	−0.051	0.025	−0.166	0.086		
	(0.109)	(0.043)	(0.028)	(0.122)	(0.100)	(0.189)		
Observations	284	243	225	200	316	213		
Bandwidth (meters)	36	28	26	27	37	41		
Panel D. False boundary 4								
RD coefficient							−0.051	−0.153
							(0.121)	(0.183)
Observations							251	83
Bandwidth (meters)							44	44

Notes: RD coefficient refers to RD coefficient from estimating equation (1) using a polynomial specification and using each of the false boundaries in each panel as the treatment boundary. Refer to Figure 5 for an illustration of these false boundaries. Clustered standard errors in parentheses. Columns 1–6 use street blocks as clusters. Columns 7 and 8 use post codes. Optimal bandwidth chosen using Calonico, Cattaneo, and Titiunik (2014). All specifications use the same baseline controls as the ones for outcomes in main analysis (Tables 2, 3, and 8). Not enough data in year 1936 to perform analysis in this table.

Table B4—Differences in Rental Price, House Occupancy Due to Cholera Exposure within BSP Area

	log rental price		Number of immigrant families at address		Proportion of immigrant families at address	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Pre-outbreak (1851, 1853)						
At least one death	−0.069		0.117		0.012	
	(0.038)		(0.091)		(0.023)	
Number of deaths		0.008		0.024		−0.000
		(0.012)		(0.029)		(0.007)
Observations	433	433	401	401	401	401
Mean	45.51	45.51	0.36	0.36	0.10	0.10
Panel B. Post-outbreak (1861, 1864)						
House has at least one death	−0.044		0.137		0.028	
	(0.036)		(0.130)		(0.043)	
Number of deaths by house		0.007		0.024		−0.003
		(0.013)		(0.046)		(0.009)
Observations	433	433	401	401	401	401
Mean	46.36	46.36	0.65	0.65	0.15	0.15

Notes: Clustered standard errors shown in parentheses. All regressions restricted to properties inside the BSP boundary. All specifications include block fixed effects. *Mean* refers to average in houses within BSP area without a death. *At least one death* is an indicator for whether a house had at least a cholera death during the 1854 outbreak. *Number of deaths* refers to the total number of cholera deaths in a house during the 1854 outbreak. *Pre-outbreak* refers to year 1853 in panel A, columns 1 and 2 and year 1851 in panel A, columns 3–6. *Post-outbreak* refers to year 1864 in panel B, columns 1 and 2 and year 1861 in panel B, columns 3–6. Immigrant families are defined by a household with a head born outside of England, Wales, or Scotland. Analysis uses block fixed effects. All specifications use baseline controls described in panels A and B of Table 3. Census data acquired from the National Archives of the UK: Public Record Office.

Table B5—Boundary Effects Using John Snow's Boundary Definition

	Pre-outbreak	Cholera exposure		Post-outbreak			
	Rental price (1853)	Number of deaths	At least one death	Rental price (1864)	Rental price (1894)	Rental price (1936)	House value (1995–2013)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Inside BSP	0.032	0.325	0.178	−0.065	−0.146	−0.239	−0.292
	(0.070)	(0.108)	(0.043)	(0.068)	(0.112)	(0.148)	(0.138)
Observations	523	500	486	602	494	181	222
Clusters	70	65	64	84	76	37	45
Bandwidth (meters)	33	27	26	42	45	31	24

Notes: All specifications use Snow (1855) definition of the catchment area boundary. *At least one death* is an indicator for whether a house had at least a cholera death during the 1854 outbreak. *Number of deaths* refers to the total number of cholera deaths in a house during the 1854 outbreak. All columns use a specification equivalent to column 3 of Table 3 (i.e., polynomial RD specification using an optimal bandwidth chosen as in Calonico, Cattaneo, and Titiunik 2014). Standard errors clustered at the block level (columns 1–6) and the post code level (column 7). All specifications use the same baseline controls as the ones for outcomes in main analysis (Tables 2, 3, and 8).

REFERENCES

Aaronson, Daniel. 2001. "Neighborhood Dynamics." *Journal of Urban Economics* 49 (1): 1–31.

Allen, Robert C. 2005. "Capital Accumulation, Technological Change, and the Distribution of Income during the British Industrial Revolution." Unpublished.

Anas, Alex. 1980. "A Model of Residential Change and Neighborhood Tipping." *Journal of Urban Economics* 7 (3): 358–70.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2004. "Tiebout Sorting, Social Multipliers and the Demand for School Quality." NBER Working Paper 10871.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy* 115 (4): 588–638.

Bayer, Patrick, Robert McMillan, and Kim Rueben. 2004. "An Equilibrium Model of Sorting in an Urban Housing Market." NBER Working Paper 10865.

Becker, Gary Stanley, and Kevin M. Murphy. 2009. *Social Economics: Market Behavior in a Social Environment.* Cambridge, MA: Harvard University Press.

Beeton, Isabella. 1861. *Mrs. Beeton's Book of Household Management.* London: S. O. Beeton.

Bergstrom, Theodore C., and Robert P. Goodman. 1973. "Private Demands for Public Goods." *American Economic Review* 63 (3): 280–96.

Bergstrom, Theodore C., Daniel L. Rubinfeld, and Perry Shapiro. 1982. "Micro-Based Estimates of Demand Functions for Local School Expenditures." *Econometrica* 50 (5): 1183–1205.

Black, Dan, Jose Galdo, and Jeffrey Smith. 2007. "Evaluating the Regression Discontinuity Design Using Experimental Data." Unpublished.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics* 114 (2): 577–99.

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

Booth, Charles. 1902. *Life and Labour of the People in London.* Vol. 1. London: Macmillan.

Brock, William A., and Steven N. Durlauf. 2001. "Discrete Choice with Social Interactions." *Review of Economic Studies* 68 (2): 235–60.

Brooks, Leah, and Byron Lutz. 2019. "Vestiges of Transit: Urban Persistence at a Microscale." *Review of Economics and Statistics* 101 (3): 385–99.

Brueckner, Jan K., Jacques-François Thisse, and Yves Zenou. 1999. "Why Is Central Paris Rich and Downtown Detroit Poor?: An Amenity-Based Theory." *European Economic Review* 43 (1): 91–107.

Calonico, Sebastian, Matias D. Cattaneo, and Rocío Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

Calonico, Sebastian, Matias D. Cattaneo, and Rocío Titiunik. 2015. "Optimal Data-Driven Regression Discontinuity Plots." *Journal of the American Statistical Association* 110 (512): 1753–69.

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

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. G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics* 92 (1): 1–45.

Conley, Timothy G. 2008. "Spatial Econometrics." Unpublished.

Dell, Melissa. 2010. "The Persistent Effects of Peru's Mining Mita." *Econometrica* 78 (6): 1863–1903.

Dericks, Gerard, and Hans R. A. Koster. 2016. "Density Restrictions and the Urban Spatial Structure: Evidence from Blitz Bombings." Unpublished.

Dowell, Stephen. 1965. *A History of Taxation and Taxes in England.* 3rd ed. London: Frank Cass & Co.

Epple, Dennis, and Holger Sieg. 1999. "Estimating Equilibrium Models of Local Jurisdictions." *Journal of Political Economy* 107 (4): 645–81.

Erwig, Martin. 2000. "The Graph Voronoi Diagram with Applications." *Networks* 36 (3): 156–63.

Farrell, Joseph, and Garth Saloner. 1985. "Standardization, Compatibility, and Innovation." *RAND Journal of Economics* 16 (1): 70–83.

Glaeser, Edward L., and Jose Scheinkman. 2000. "Non-Market Interactions." Unpublished.

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

Hanlon, W. Walker. 2017. "Temporary Shocks and Persistent Effects in Urban Economies: Evidence from British Cities after the U.S. Civil War." *Review of Economics and Statistics* 99 (1): 67–79.

Heblich, Stephan, Alex Trew, and Yanos Zylberberg. 2016. "East Side Story: Historical Pollution and Persistent Neighborhood Sorting." Unpublished.

Holmes, Thomas J. 1998. "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders." *Journal of Political Economy* 106 (4): 667–705.

Hornbeck, Richard, and Daniel Keniston. 2017. "Creative Destruction: Barriers to Urban Growth and the Great Boston Fire of 1872." *American Economic Review* 107 (6): 1365–98.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–35.

Ioannides, Yannis Menelaos. 2013. *From Neighborhoods to Nations: The Economics of Social Interactions.* Princeton, NJ: Princeton University Press.

Jackson, Kenneth T. 1987. *Crabgrass Frontier: The Suburbanization of the United States.* New York: Oxford University Press.

Johnson, Steven. 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.

Kain, John F., and John M. Quigley. 1975. "Introduction." In *Housing Markets and Racial Discrimination: A Microeconomic Analysis,* 1–8. New York: NBER.

Kane, Thomas J., Stephanie K. Riegg, and Douglas O. Staiger. 2006. "School Quality, Neighborhoods, and Housing Prices." *American Law and Economic Review* 8 (2): 183–212.

Keele, Luke J., and Rocío Titiunik. 2015. "Geographic Boundaries as Regression Discontinuities." *Political Analysis* 23 (1): 127–55.

Kremer, Michael, Jessica Leino, Edward Miguel, and Alix Peterson Zwane. 2011. "Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions." *Quarterly Journal of Economics* 126 (1): 145–205.

Lagunoff, Roger, and Akihiko Matsui. 1997. "Asynchronous Choice in Repeated Coordination Games." *Econometrica* 65 (6): 1467–77.

Lalive, Rafael. 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 (accessed December 19, 2014).

Lee, David S. 2008. "Randomized Experiments from Non-Random Selection in U.S. House Elections." *Journal of Econometrics* 142 (2): 675–97.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281–355.

Lee, Sanghoon, and Jeffrey Lin. 2018. "Natural Amenities, Neighbourhood Dynamics, and Persistence in the Spatial Distribution of Income." *Review of Economic Studies* 85 (1): 663–94.

LonRes. 2015. "LonRes: Rental Price Archives." Access provided by Greater London Properties. http://www.lonres.com/public/ (accessed June 20, 2015).

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60 (3): 531–42.

Manski, Charles F. 2000. "Economic Analysis of Social Interactions." *Journal of Economic Perspectives* 14 (3): 115–36.

Maskin, Eric, and Jean Tirole. 1988a. "A Theory of Dynamic Oligopoly I: Overview and Quantity Competition with Large Fixed Costs." *Econometrica* 56 (3): 549–69.

Maskin, Eric, and Jean Tirole. 1988b. "A Theory of Dynamic Oligopoly II: Price Competition, Kinked Demand Curves, and Education Cycles." *Econometrica* 56 (3): 571–99.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2): 698–714.

Mieszkowski, Peter, and Edwin S. Mills. 1993. "The Causes of Metropolitan Suburbanization." *Journal of Economic Perspectives* 7 (3): 135–47.

Möbius, Markus M. 2000. "The Formation of Ghettos as a Local Interaction Phenomenon." Unpublished.

Muth, Richard F. 1969. *Cities and Housing; The Spatial Pattern of Urban Residential Land Use.* Chicago: University of Chicago Press.

Oates, Wallace E. 1969. "The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis." *Journal of Political Economy* 77 (6): 957–71.

O'Day, Rosemary, and David Englander. 1993. *Mr. Charles Booth's Inquiry: Life and Labour of the People in London Reconsidered.* London: Hambledon Press.

Okabe, Atsuyuki, Barry Boots, Kokichi Sugihara, Sung Nok Chiu, and D. G. Kendall. 2000. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams.* 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc.

Ordnance Survey. 1951. *Ordnance Survey: National Grid Maps, 1940s–1960s.* Edinburgh: National Library of Scotland.

Pancs, Romans, and Nicolaas J. Vriend. 2007. "Schelling's Spatial Proximity Model of Segregation Revisited." *Journal of Public Economics* 91 (1–2): 1–24.

Quigley, John M. 1974. "Racial Discrimination and Housing Consumption of Black Households." In *Patterns of Racial Segregation, Vol. 1: Housing,* edited by George M. von Furstenberg, Bennett Harrison, and Ann R. Horowitz. Lexington, MA: Lexington Books.

Schelling, Thomas C. 1969. "Models of Segregation." *American Economic Review* 59 (2): 488–93.

Schelling, Thomas C. 1971. "Dynamic Models of Segregation." *Journal of Mathematical Sociology* 1 (2): 143–86.

Schelling, Thomas C. 1978. *Micromotives and Macrobehavior.* New York: W. W. Norton and Co.

Sherwell, Arthur. 1897. *Life in West London.* London: Methuen & Co.

Shiode, Shino. 2012. "Revisiting John Snow's Map: Network-Based Spatial Demarcation of Cholera Area." *International Journal of Geographical Information Science* 26 (1): 133–50.

Siodla, James. 2015. "Razing San Francisco: The 1906 Disaster as a Natural Experiment in Urban Redevelopment." *Journal of Urban Economics* 89: 48–61.

Snow, John. 1855. *On the Mode of Communication of Cholera.* London: John Churchill.

Summers, Judith. 1991. *Soho: A History of London's Most Colourful Neighbourhood.* London: Bloomsbury Publishing.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64 (5): 416–24.

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

Whitehead, Richard. 1854. *The Cholera in Berwick Street. The Senior Curate of St. Luke's.* London: Hope & Co.

Zoopla. 2015. "Zoopla Current Value Estimates." http://www.zoopla.co.uk (accessed February 20, 2015).

This article has been cited by:

1. Hans R.A. Koster, Edward W. Pinchbeck. 2022. How Do Households Value the Future? Evidence from Property Taxes. *American Economic Journal: Economic Policy* **14**:1, 207-239. [Abstract] [View PDF article] [PDF with links]
2. Haiqiang Chen, Wenlan Qian, Qiang Wen. 2021. The Impact of the COVID-19 Pandemic on Consumption: Learning from High-Frequency Transaction Data. *AEA Papers and Proceedings* **111**, 307-311. [Abstract] [View PDF article] [PDF with links]