Segregation and the Initial Provision of Water in the United States*

Brian Beach
William & Mary
and NBER

John Parman
William & Mary
and NBER

Martin Saavedra
Oberlin College

December 17, 2018

Abstract

We examine the extent to which segregation shaped the initial provision of water in the United States. We develop a theoretical model to illustrate how segregation affects the extensiveness of water systems. Data from over 1700 cities and towns match the key empirical predictions of our model: waterworks were built earlier in larger and more segregated cities as well as cities with smaller black shares. In the context of our model, these results are consistent with blacks in segregated cities being excluded from water provision. Analysis of health outcomes further supports this interpretation. Segregated cities experienced smaller health improvements following the construction of a waterworks and were much slower to eliminate waterborne diseases. This suggests that, by facilitating the exclusion of black households, segregation also served to undermine the city's ability to eliminate waterborne diseases for all residents.

---

[*] Contact information for Beach: bbbeach@wm.edu; Parman: jmparman@wm.edu; Saavedra: martin.saavedra@oberlin.edu. Thanks to Francisca Antman and seminar participants at George Mason University, the Western Economic Association's annual meeting, the American Historical Association's annual meeting, and the World Economic History Congress.

1 Introduction

At the dawn of the 20th century, mortality rates in the United States were much higher in cities than in rural areas. This 'urban mortality penalty' was a common feature among industrial nations during this period (Cain & Hong, 2009; Kesztenbaum & Rosenthal, 2011). As to the causes of this penalty, the literature has settled on three factors: infectious diseases, particularly those associated with unclean water and improper sewage disposal, poor nutrition, and large amounts of air pollution from the burning of coal.[1] Although the precise contribution of each channel remains an open question, the existing literature suggests that investments in water and sewerage played an important role in reducing mortality in the United States and elsewhere. For instance, Alsan & Goldin (Forthcoming) study 60 Massachusetts municipalities and estimate that clean water and effective sewerage systems account for roughly one-third of the decline in infant mortality between 1880 and 1920.[2]

While its results are impressive, this public health movement would not have been possible without earlier infrastructure investment. As Figure 1 illustrates, U.S. cities built waterworks at a tremendous rate during the 1890s: by 1897 there were 3,167 waterworks in the United States with 83% built after 1880 and 46% built between 1890 and 1896.[3] These investments, which occurred before effective water treatment

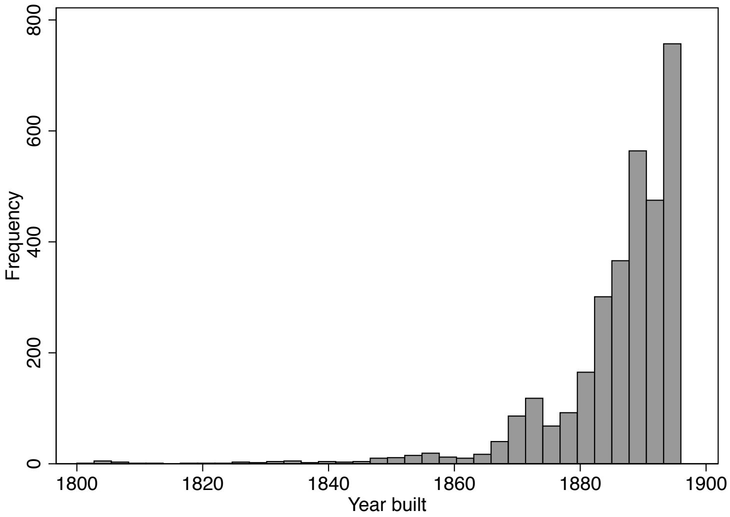
---

[1] On the role of waterborne diseases see (Ferrie & Troesken, 2008; Alsan & Goldin, Forthcoming; Antman, 2015). For more on poor nutrition see (McKeown, 1976; Fogel, 2004; Fogel & Costa, 1997). For air pollution see (Beach & Hanlon, 2018; Clay *et al.* , 2015).

[2] Ferrie & Troesken (2008) document similar patterns in their case study of Chicago, but they also document a significant decline in mortality for those over the age of five as well. Recent work by Kesztenbaum & Rosenthal (2017) examines the diffusion of sewerage though the neighborhoods of Paris and also finds meaningful declines in mortality risk across the age distribution. Others have also documented a strong link between increased infrastructure expenditure and subsequent mortality declines (e.g., Cain & Rotella (2001) and Chapman (2018)). An alternative approach is to examine specific interventions within a differences-in-differences framework. An influential paper using this approach is Cutler & Miller (2005), which examines the experience of 13 U.S. cities and argues that roughly half of that decline can be explained by the introduction of clean water technologies. Recent work by Anderson *et al.* (2018) suggests the magnitude is probably closer to 10%, though they still document a strong link between these technologies and the elimination of typhoid fever.

[3] This statistic based on data from Baker (1897), the most comprehensive source on U.S. water-

Figure 1: Distribution of waterworks construction dates



Data from Baker (1897). Sample includes 3,167 waterworks built between 1800 and 1897.

technologies existed, laid the necessary foundation for the elimination of infectious waterborne diseases by connecting households to a centralized water supply. Without these connections, future investments in water filtration would have been much less effective at combatting waterborne disease, as residents with access to clean water would remain at risk through their contact with residents without access.[4]

Given how closely these two movements are linked, it is surprising how little we know about the factors shaping this initial wave of investment. Why American cities decided to build when they did and what determined the extensiveness of the system they built are both open questions. These questions are also of central importance for understanding the broader success of the movement to filter water and treat sewage.

This paper examines the extent to which the segregation of black households in-

---

works. Section 3.1 describes this source in more detail.

[4] The waterborne disease typhoid fever, for instance, is typically spread through the consumption of water that is tainted from the fecal waste of infected individuals. However, infection from other forms of contact is also possible. The most well-known alternative transmission is probably "Typhoid Mary", an asymptomatic carrier of typhoid who is presumed to have infected over 50 people with typhoid fever during her career as a cook.

fluenced the timing of investment and the overall extensiveness of the system. Our motivation for considering the role of segregation is twofold. First, building a waterworks was incredibly costly: in 1890, the median waterworks cost over $100,000 or $25 per-capita. In today's dollars a comparable project would cost $561 per-capita.[5] Since these projects were financed at the local level, investment likely depended on the political will of local residents. This brings us to our second motivation. Since these investments were made before effective water filtration and sewage treatment existed, taxpayers had little incentive to attempt to internalize the future social returns that would be generated from universal access.[6] When these two facts are considered together, it becomes clear that segregation may have incentivized cities to exclude lower status groups as a way of lowering the cost of provision.

We develop a theoretical model to illustrate how segregation might influence water provision. We then draw on full count census data from 1880 to quantify the degree of segregation in all incorporated U.S. cities.[7] We pair this information with data that we digitize from Baker (1897) – a comprehensive source on the history of American waterworks. These two sources allow us to test the predictions of our model. Consistent with these predictions, we find that segregated cities were quicker to build a waterworks.

While the model and year built results are consistent with an exclusion story, we draw additional support for this interpretation from our analysis of health data. If segregated cities excluded black households from these initial investments, then this should undermine the city's ability to combat infectious disease. Because comprehen-

---

[5] Median waterworks cost is available from Baker (1897), although the figure we quote is based on the incorporated cities that we use for our analysis. Our price adjustment comes from Measuring Worth.

[6] Cutler & Miller (2005) estimate the social rate of return on clean water technologies was greater than 23 to 1. This return only considers life expectancy gains, however, Beach *et al.* (2016) show that early-life exposure to waterborne disease impaired human capital development, affecting earnings and educational attainment. By helping eliminate this exposure, Beach *et al.* estimate that the gains to future income alone offset cost of building a waterworks by a factor of 4 to 1.

[7] Our measure of segregation follows (Logan & Parman, 2017) but is ultimately calculated at the city rather than the county level. This measure is discussed in more detail in Section 3.2.

sive city-level mortality statistics were not collected before 1900, we use information from the 1900 census to generate a proxy for infant mortality. In that census all women were asked the number of children they have given birth to as well as how many of those children were still alive. We use this information to generate an indicator for whether a mother has ever lost a child, which we interpret as a proxy for infant mortality.[8] We then adopt a difference-in-differences empirical specification that exploits variation across cities in the timing of construction as well as variation in the number of fertile years that women within the same city spent exposed to that waterworks.

Results from this specification indicate that white mothers were less likely to lose a child following the construction of a waterworks and this effect does not vary with the degree of segreation in the city. For black mothers, however, the declines in infant mortality were particularly pronounced in integrated cites, and the benefits diminished as the level of segregation increased. This is consistent with the idea that black households in integrated cities gained access to these new water systems due to living in close proximity to white households; conditional on supplying a white household with water, the marginal cost of supplying a nearby black household with water was very low. This interpretation is further supported by annual mortality data. We transcribe annual city-level mortality data from 1900-1930 and examine whether segregated cities were slower to eliminate waterborne diseases like typhoid fever. While all cities in our sample eliminated typhoid fever by 1930, for any given threshold, segregated cities were always slower to lower typhoid fever rates below that threshold, further illustrating the lasting consequences of those initial provision decisions.

These results complement previous work by Troesken (2002), which we extend in several ways. Troesken posits that controlling waterborne disease requires compre-

---

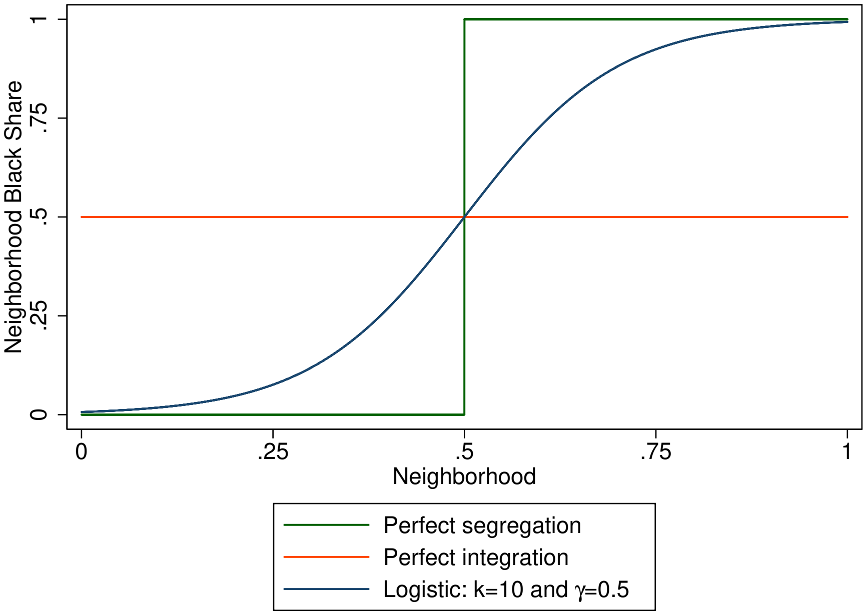
[8] The dramatic gains in life expectancy that occurred from 1880 to 1950 were largely the result of decreasing mortality for those under the age of five, see Preston & Haines (1991). Life expectancy conditional on surviving to age 20 did not change much during this period.

hensive access, which is more likely in an integrated city rather than a segregated city. Troesken's analysis, however, largely centers on assessing the efficacy of the adoption of clean water technologies. In contrast, our paper is much more concerned with assessing the robustness of the link between segregation and the extensiveness of a city's waterworks. To this end, we formalize Troesken's intuition in order to generate a set of empirical predictions, which we then test empirically. A second point of differentiation is that, as a result of recently digitized data, we extend the scope of this research to far more cities than previously possible. The recent release of the complete count federal census data allows us to construct historical segregation estimates for every city in the United States (Logan & Parman, 2017). Troesken, in contrast, was forced to rely on a sample of 33 large cities because of the availability of both segregation and mortality data. Our sample of over 1700 cities captures a much broader range of city sizes, population compositions, and political environments. Despite all of these differences, we find robust support for Troesken's original hypothesis. Our findings illustrate that the ideas presented in Troesken (2002) are much more generalizable than the empirical support in Troesken's paper would otherwise suggest.

2 Modeling the influence of segregation on local water provision

Consider a city with two types of residents. For ease of exposition, we define the two groups as white and black; however, we could consider any minority group that potentially faces discrimination in the provision of public goods. The city lies on a unit interval with $x = 0$ being the whitest part of town and $x = 1$ being the blackest. Without loss of generality, assume the city is one mile long. Each point along the line is a neighborhood. The degree of segregation and the group sizes are characterized by an increasing function $g()$, which indicates the proportion of the neighborhood that

Figure 2: Examples of $g(.)$



Notes: Each line corresponds to the distribution of neighborhoods for a hypothetical city with a total black share of 50%. Neighborhoods are organized based on their black share with 0 being the neighborhood with the smallest black share and 1 being the neighborhood with the largest black share. All neighborhoods are assumed to be of the same size.

is black. For example, if the city is perfectly segregated and each group makes up one half of the city, then:

$$g(x) = \begin{cases} 0 & \text{if } x \leq \frac{1}{2}; \\ 1 & \text{if } x > \frac{1}{2}. \end{cases}$$

On the other hand, perfect integration for a city with equal group sizes implies that $g(x) = \frac{1}{2}$. More typically, we would expect $g()$ to be an increasing S-shaped function. Figure 1 graphically depicts $g()$ for our hypothetical city under perfect segregation, perfect integration, and something in between. The actual $g()$ functions for several of the cities in our sample are provided in Appendix A.

Now suppose that city politicians face the budget constraint $B = z + c * m + FC$, where B is the city's budget, z is non-water related public goods with a price normalized to 1, m are miles of water mains, c is the per mile cost of water mains, and FC represents any fixed costs associated with supplying water. If the city does not build any mains, then the constraint is $B = z$.

City officials value non-water related public goods, white residents with access to clean water, and black residents with access to clean water. City officials will place lower value on black resident access to water, which may reflect taste-based or statistical discrimination (e.g., the fact that black residents are more likely to be disenfranchised and thus less likely to help keep city officials in office). Thus, if the city builds a main, it will start at $x = 0$ (the whitest part of town), and keep building the main, possibly stopping before supplying water to the whole city. If a main is built to a particular neighborhood x, then both black and white residents of that neighborhood have access to the main. Now let N_W be the white population that is connected to a water main, N_B be the black population connected to a water main, and $m \in [0, 1]$ be where the city stops building the main. The variable m is the miles of mains and a measure of the water system's size. Then

$$N_W = \int_0^m (1 - g(x))dx \tag{1}$$

$$N_B = \int_0^m g(x)dx \tag{2}$$

Now suppose the city's utility function is:

$$U(N_W, N_B, Z) = \alpha N_W + (1 - \alpha)N_B + \beta z \tag{3}$$

where $\alpha \in (\frac{1}{2}, 1)$. The utility function in (N_w, N_b, z)-space is linear, but since g is nonlinear, the utility function in (z, m)-space is non-linear.

For an interior solution, we need the ratio of marginal utilities of m and z to be equal to the ratio of the costs of m and z. Since $\frac{1-\alpha}{\beta} \leq \frac{MU_m}{MU_z} \leq \frac{\alpha}{\beta}$, an interior solution will require that the per-mile cost $c \in \left(\frac{1-\alpha}{\beta}, \frac{\alpha}{\beta}\right)$. If $c \leq \frac{1-\alpha}{\beta}$, then the city will provide water to all residents. If $c \geq \frac{\alpha}{\beta}$, then the city will not provide water to any residents.

For an interior solution, taking the first-order conditions indicates that:

$$m^* = g^{-1}\left(\frac{\alpha - \beta c}{2\alpha - 1}\right) \tag{4}$$

Since g is an increasing function, it follows that g^{-1} is an increasing function. Let $\lambda = \frac{\alpha-\beta c}{2\alpha-1}$. Then $\frac{\partial\lambda}{\partial c} = \frac{-\beta}{2\alpha-1} < 0$. Thus, an increase in the cost of a water main decreases optimal main mileage. Similarly, an increase in the preferences for non-water public goods decreases optimal main mileage. As for the preferences for whites, $\frac{\partial\lambda}{\partial\alpha} = \frac{2\beta c-1}{(1-2\alpha)^2}$, which is theoretically ambiguous.

Next, let's characterize the function g to analyze the effects of segregation and group size. Let

$$g(x) = \frac{1}{1 + e^{-k(x-\gamma)}} \tag{5}$$

This is an S-shaped curve in which k measures the degree of segregation. As k goes to infinity, the city becomes perfectly segregated, while $k = 0$ implies perfect integration. The parameter γ (the centering parameter) is the proportion of the city that is white. Then,

$$g^{-1}(x) = \gamma - \frac{1}{k}\ln\left(\frac{1}{x} - 1\right). \tag{6}$$

This implies that m^* increases as white share increases. The effect of segregation on water provision is ambiguous. If $\beta c > \frac{1}{2}$, then segregation increases optimal main mileage. If $\beta c < \frac{1}{2}$, then segregation decreases optimal main mileage.

Proposition 2.1 *If the optimal main stops in a neighborhood that is less than one half black, then a marginal increase in either segregation or the preference for whites increases the size of the optimal water system. If the optimal main stops in a neighborhood that is greater than one half black, then a marginal increase in either segregation or the preference for whites decreases the size of the optimal water system.*

Proof A marginal increase in the preference for whites or a marginal increase in segregation increases the size of the water system if and only if $\beta c > \frac{1}{2}$. This implies that

$$\frac{\alpha - \beta c}{2\alpha - 1} < \frac{\alpha - \frac{1}{2}}{2\alpha - 1} < \frac{1}{2}. \tag{7}$$

Since g^{-1} is an increasing function, this implies that $g^{-1}\left(\frac{\alpha-\beta c}{2\alpha-1}\right) < g^{-1}\left(\frac{1}{2}\right)$. Applying g to both sides of this inequality yields the result.

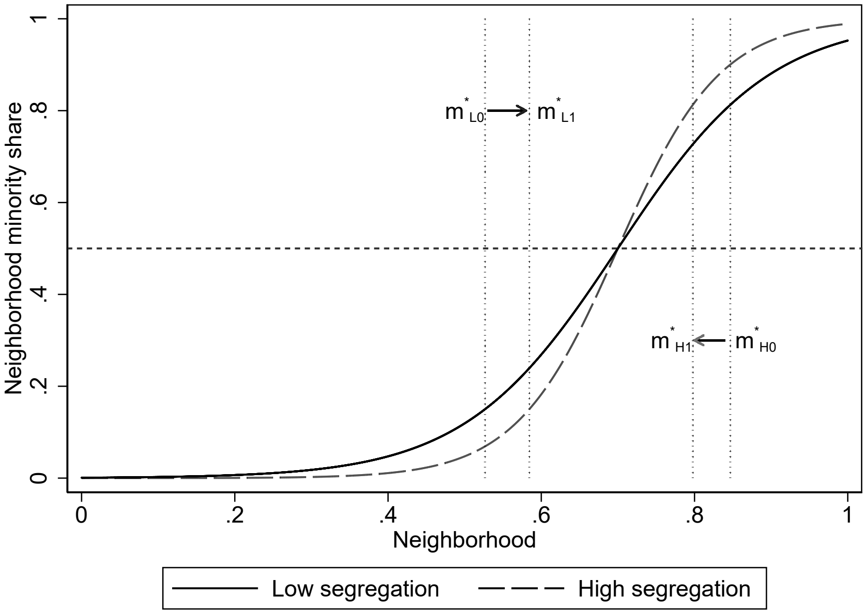
Intuitively, more segregation decreases the marginal benefit from continuing construction only if the city is already building in a minority neighborhood. If the city has not yet completed construction in the majority neighborhood, more segregation implies there is a higher utility gain to the marginal water main mile. This result is displayed visually in Figure 3, which shows the effects of an increase in segregation on m^*. The vertical line m^*_{L0} shows an optimal main that stops in a majority white neighborhood. After the increase in segregation, the optimal main increases in length to m^*_{L1}. The vertical line m^*_{H0} shows an optimal main for the original level of segregation, but because of different values of α, β, and c, the optimal main stops in a majority black neighborhood. After the increase in segregation, the optimal main decreases to m^*_{H1}.

These results all pertain to a city at an interior solution making decisions on the intensive margin of how many miles of mains to build. However, equally important are city decisions on the extensive margin of whether to build a waterworks at all. Given relatively low marginal costs to extending the water system, c, the sign of the marginal effects of an increase in segregation or in preference for whites will be the same for the likelihood of building a system as for changing the size of a system.

Proposition 2.2 *If the marginal costs of extending a water system are sufficiently small, the sign of the marginal effect of an increase in either segregation or preference for whites on the likelihood of constructing a waterworks is the same as the sign of the marginal effect on the optimal size of a water system.*

Proof If the city does not build a water water their utility is simply $U(\neg\text{Build}) =$

Figure 3: Comparative statics of how changes in segregation influence main provision



Notes: The white share for this hypothetical city is set to 0.7. In both segregation scenarios, $g()$ is modeled as a logistic function. For the low segregation scenario we set k to 10 and in the high segregation scenario we set k to 15. The optimal mains to the left of 0.7 are calculated assuming $\beta = 0.8; c = 0.8$; and $\alpha = 0.7$. The optimal mains to the right of 0.7 are calculated assuming $\beta = 0.5; c = 0.5$; and $\alpha = 0.9$

βB. If the city does build a waterworks, they consume on the budget constraint $B = z + c \times m^* + FC$. Plugging this into their utility function gives

$$U(\text{Build}) = \int_0^{m^*} (\alpha(1 - g(x)) + (1 - \alpha)g(x))\, dx + \beta(B - c \times m^* - FC) \tag{8}$$

Thus, the city will build so long as

$$U(\text{Build}) - U(\neg\text{Build}) = \int_0^{m^*} (\alpha(1 - g(x)) + (1 - \alpha)g(x))\, dx - \beta(c \times m^* + FC) > 0 \tag{9}$$

Taking the derivative with respect to m^* yields $\alpha + (1 - 2\alpha)g(m^*) - \beta c$, giving us the marginal effect of an increase in the optimal water system size on the gain in utility from building the system. Thus, if $\frac{1}{\beta}(\alpha + (1 - 2\alpha)g(m^*)) > c$, in other words so long as marginal costs are sufficiently low.

The intuition behind this is that if the optimal system size is large, the cost of building it is more likely to cover the fixed cost, but only if the marginal cost are not excessively high. Thus if a change in segregation or white preferences increases the optimal system size conditional on having a waterworks, that change will also increase the likelihood that building the system overall is optimal. As shown earlier, if the optimal stopping point is in a mostly white neighborhood, an increase in segregation will increase the size of the waterworks. Thus, changes in segregation will have an analogous effect on year built: if the marginal cost is sufficiently low, an increase in segregation should push a city to build earlier.

Thus far we have considered a model where city officials place lower weight on black access to water. This is reflected in the preference parameter α, which can be thought of as capturing local officials' need to cater to likely voters, desire to focus provision of public goods on those most able to finance those goods, or discriminatory preferences for a particular group. Since roughly half of local waterworks were privately owned during this period, a natural question is whether private provision affects the predictions of our model. As private firms should be motivated by profits, one would expect that private firms would not face the same pressure to discriminate. In practice, however, private firms would still likely treat potential white and black customers differently. In particular, if white residents have higher ability or willingness to pay for water service on average, the firm's profit function will effectively place greater weight on neighborhoods with more white residents as those neighborhoods would contain more paying customers and could thus be served with far lower average costs per customer. Accordingly, private provision does not affect the comparative statics of our model. These results are available in Appendix B.

3 Taking the model to the data

Our model of water provision leaves us with several testable predictions. First, the size of the optimal water system increases with the size of the majority group. Second, in a city in which the last water main stops in a neighborhood less than one half minority, an increase in segregation should lead to an increase in the size of the optimal water system. Third, in a city in which the last water main stops in a neighborhood more than one half minority, an increase in segregation should lead to a decrease in the size of the optimal water system. Finally, we have an analogous set of predictions for the timing of the initial decision to build a water system: cities should establish water systems earlier if the size of the majority group increases or if segregation increases (decreases) in a city in which the last water main stops in a neighborhood less than (more than) one half minority.

The ideal dataset to examine these predictions would include: city-level variation in black share, city-level variation in the segregation of the black population, and high quality data that allows us to identify access to water and racial composition at the neighborhood level. Our primary data constraint is race-tabulated water provision at the household or even neighborhood level as well as exogenous variation in the black population share as well as its spatial distribution. For this reason, our primary results are going to focus on our final empirical prediction – that segregated cities are likely to build their waterworks earlier – and then we will examine several proxies for access, which offers empirical support for our hypothesized mechanism.

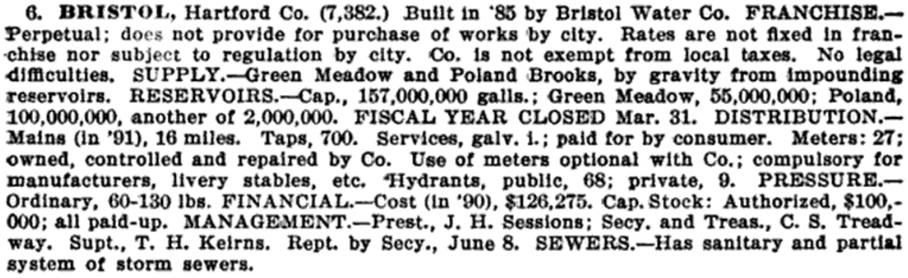
3.1 Water provision data

Our primary source on water provision is the 1897 edition of Moses Nelson Baker's *Manual of American Waterworks*. Baker's first volume of the Manual of American Waterworks was published in 1888. As is made clear from his introduction, Baker's efforts were inspired by his frustration with the fact that essential data on America's waterworks (e.g., the number of waterworks, the location of each waterworks, and when each waterworks was built) were not accurately known. Baker attempted to remedy this situation by surveying local officials and companies to obtain accurate histories and continuing to follow up with these individuals until he received the requested information. Subsequent editions focused on standardizing the information obtained from those surveys so that the information could be included in future editions. Baker also incorporated information on works built since the initial survey. For these reasons, we focus on digitizing the 1897 edition of the manual as it is the most comprehensive of the four volumes.

For each waterworks, the manual includes information on "its history, general character, the capacity of the pumps, reservoirs, stand-pipes, or filters; the extent of the distribution system; and the most important figures relating to the finances of each system" (Baker (1897), preface). A representative example of one these descriptions is provided in Figure 4. The main piece of information we focus on is the year the waterworks was built. We digitize this information for all 4,207 cities appearing in the 1897 manual.

It is worth pointing out why we prefer to use year built instead of other measures,

Figure 4: Example entry from Baker (1897)



6. **BRISTOL,** Hartford Co. (7,382.) Built in '85 by Bristol Water Co. FRANCHISE.—Perpetual; does not provide for purchase of works by city. Rates are not fixed in franchise nor subject to regulation by city. Co. is not exempt from local taxes. No legal difficulties. SUPPLY.—Green Meadow and Poland Brooks, by gravity from impounding reservoirs. RESERVOIRS.—Cap., 157,000,000 galls.; Green Meadow, 55,000,000; Poland, 100,000,000, another of 2,000,000. FISCAL YEAR CLOSED Mar. 31. DISTRIBUTION.—Mains (in '91), 16 miles. Taps, 700. Services, galv. i.; paid for by consumer. Meters: 27; owned, controlled and repaired by Co. Use of meters optional with Co.; compulsory for manufacturers, livery stables, etc. Hydrants, public, 68; private, 9. PRESSURE.—Ordinary, 60-130 lbs. FINANCIAL.—Cost (in '90), $126,275. Cap. Stock: Authorized, $100,000; all paid-up. MANAGEMENT.—Prest., J. H. Sessions; Secy. and Treas., C. S. Treadway. Supt., T. H. Keirns. Rept. by Secy., June 8. SEWERS.—Has sanitary and partial system of storm sewers.

such as miles of mains or number of taps, both of which are observed in the Baker data and relate directly to the intensive margin of water system construction considered in our model. The model uses miles of mains to illustrate which neighborhoods will get access and which neighborhoods will not. The prediction of the model is that, depending on the demographic mix of the final neighborhood that receives water, an increase in segregation can either increase or decrease the extensiveness of the system. As shown earlier, when the marginal neighborhood is majority black then an increase in segregation will decrease water provision. Conversely, we saw that when the marginal neighborhood was mostly white an increase in segregation would increase the extensiveness of the waterworks. Unfortunately, when miles of mains are reported at the city-level it is not possible to identify the demographic mix of the final neighborhood that receives water. This is problematic because if we don't observe the demographic mix of the final neighborhood receiving water then we don't have a clear empirical prediction that can be validated with the data.

Further complicating the interpretation of a measure of extensiveness such as miles of mains is the difficulty in comparing the variable across cities. Cities in which the neighborhoods with large white population shares are located geographically close to one another may be connected to the waterworks with a small number of miles of mains relative to cities in which these same neighborhoods are spread out. Normalizations such as miles of mains per acre may help improve comparability across cities but will inevitably fail to capture certain aspects of the geography of water systems relevant to estimating the marginal effect of segregation on water system size.

To overcome these issues we note that the model does offer clear predictions for how segregation should influence the city's decision of when to build. Because the construction of a waterworks involves large fixed costs, the rational city will only begin construction when there are enough neighborhoods to justify the initial investment. That is, the first wave of construction has a marginal cost that is equal to the initial fixed costs as well as the variable costs of supplying each neighborhood (up until the optimal stopping point). As a city becomes more segregated, then all else equal there should be more neighborhoods in that initial wave of construction to spread those fixed costs across. As a result, segregated cities should begin construction earlier than integrated cities.

3.2 Demographic data

With our empirical prediction in hand, we now attach demographic characteristics to our waterworks data. Our demographic data come from the complete count 1880 federal census as maintained by the Integrated Public Use Microdata Series (IPUMS) (Ruggles *et al.* , 2015). We rely on this dataset to identify: number of households residing in the city, the black share, and the level of segregation in the city. We begin with all individuals residing in an incorporated place in 1880. By using incorporated places we are not restricted to the sample of large cities that are typically reported in census publications but are instead able to obtain demographic data for roughly 5,400 cities and villages. While one would ideally have observations of demographics for these incorporated places over time, data limitations restrict us to the 1880 census. As discussed in King & Magnuson (1995), the 1880 census presented a departure from prior censuses in terms of the approach to enumeration. Concerted efforts were made to reduce undercounts, making the census far more reliable than earlier censuses for our purposes. Unfortunately, the original manuscripts of the 1890 federal census were destroyed, making it impossible to get the complete count data needed for our purposes. We could turn to the 1900 federal census but this took place well after the waterworks in our study were built and would therefore capture demographic patterns that are likely to be driven by responses to waterworks construction rather than the drivers of that construction.

We measure segregation using the neighbor-based segregation index developed by Logan & Parman (2017). This measure exploits the public availability of complete census returns for any census over 72 years old and the door-to-door enumeration process by which census information was collected. Given that enumerators visited each household in sequential order down a street, the households appearing before and after an individual on the census manuscript page correspond to that individual's next-door neighbors. It is therefore possible to see how often individuals live next to a person of a different race, providing a very simple and intuitive way to think about the level of residential integration in a community.

The neighbor-based segregation index compares the number of black households in a given area living next to white neighbors to the number expected under complete integration and under complete segregation. Formally, the index is given by

$$\text{Neighbor-based segregation} = \frac{E(\overline{x_{Min}}) - x_{Min}}{E(\overline{x_{Min}}) - E(\underline{x_{Min}})} \tag{10}$$

where x_{Min} is the number of minority households with majority neighbors, $E(\overline{x_{Min}})$ is the expected number of minority households with majority neighbors under complete integration (the group membership of neighbors are completely independent) and $E(\underline{x_{Min}})$ is the expected number of minority households with majority neighbors under complete segregation (only the minority households on either end of the minority neighborhood have majority neighbors). This index equals zero under complete integration, increases as the number of minority households with a majority neighbor decreases, and equals one in the case of complete segregation. This measure is particularly appealing in our context since it matches the relevant geography for water provision: street segments.

Because we are examining cities and villages before the Great Migration there are a number of cities in our sample that don't have meaningful black population sizes. We restrict our sample to cities with at least 11 black households, which corresponds to the 50th percentile of all incorporated places that have at least 1 black household. This leaves us with a sample of 1,754 cities and villages, 876 of which appear in the Baker Manual. Because of the comprehensiveness of the Baker manual, we classify the remaining 878 cities and villages as not having a waterworks. Some of these waterworks could simply be non-responders, but Baker suggests the non-response share is actually quite small.

One concern with our approach to segregation is that we rely on variation across cities at a single point at time. Much of this variation will be driven by unobserved differences across cities in racial attitudes and the evolution of those attitudes, differences that will likely directly impact discrimination in public goods provision. Any estimated effects of segregation on the provision of water may be driven by the direct effect of segregation as shown in our model or by an indirect effect driven by these underlying racial tensions that gave rise to residential segregation in the first place. These differences can be thought of as more closely related to the racial preference parameter in the theoretical model. However, recall that an increase in segregation and an increase in the preference for whites both impact the optimal water system in the same direction. Thus our estimates can be thought of as capturing these additive impacts of the direct effect of segregation and the indirect effects of the unobserved racial dynamics leading to a city's current level of segregation.

4 Main results

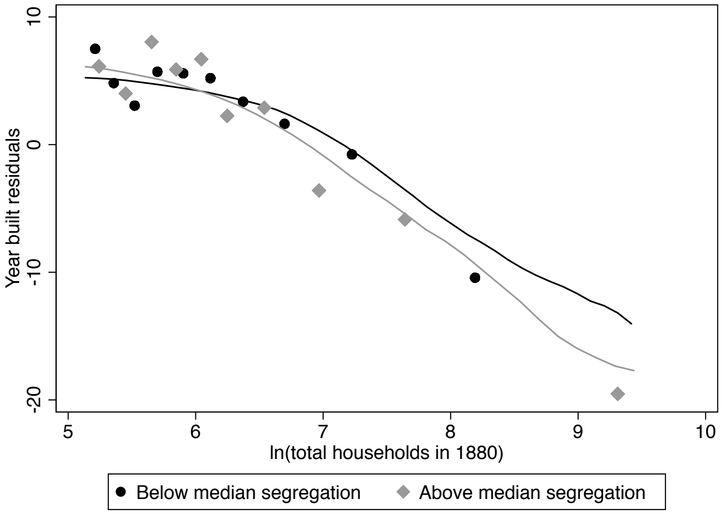
4.1 Segregation and water provision

Figure 5 visually displays our first result. We begin by regressing year built on a series of census region indicators to tease out geographic variation in year built and black shares. By definition, this sample is restricted to cities that build a waterworks before 1897. From that regression we take the residuals and plot them against ln(number of households in the incorporated place). We then fit two separate local polynomial regressions, one for the sample of cities with above median segregation and one for the sample of cities with below median segregation. Finally, we include a binned scatter plot for these two groups to illustrate the overall fit of our polynomials.

There are three main takeaways from Figure 5. First, we see a clear downward trend, which illustrates that larger cities tended to build earlier than smaller cities. This is not surprising given the inherent fixed costs associated with building a waterworks and the fact that larger cities will be able to spread those costs over more households. The second takeaway, however, is that in the largest cities, those with higher levels segregation tend to build their waterworks about 5 years earlier than cities with lower levels of segregation. This is consistent with the empirical predictions of the model outlined above. Finally we see this difference converge towards zero as the city size decreases. We hesitate to over interpret what convergence might mean given the simplicity of this specification, but it could suggest that systematic exclusion of blacks from water provision was only possible in larger cities.

Next we incorporate information from cities that did not build a waterworks by 1897 and estimate a cox proportional hazard model, taking year of construction as our "failure." These results are presented in Table 1. Column 1 examines the full sample and indicates that segregated cities built waterworks at a rate about 10 percent higher than integrated cities. We also see that cities with larger black shares are built at a

Figure 5: Non-parametric estimates of the relationship between city size and year built for high and low segregation cities



Year build residuals obtained by regressing year built on a series of census region fixed effects. The median level of segregation is 0.28.

much smaller rate. In columns 2 through 5 we restrict the sample based on city size. These results indicate that the impact of segregation is most pronounced in larger cities. For cities that fall in the 80 to 95th percentiles (so 995 to 4474 households) we see that segregated cities built at a rate roughly 5 times higher than integrated cities. In the top 5th percent of the city size distribution segregated cities built at a rate 34 times faster than integrated cities.

4.2 Did segregation facilitate excluding black households?

The above results show that segregated cities tended to build their waterworks before integrated cities. This result is consistent with our theoretical predictions, but to increase our confidence in interpreting this relationship as causal, it is worth assessing whether black households in segregated cities were less likely to have access to water infrastructure. If so, then this would offer evidence consistent with the mechanisms

Table 1: Examining the relationship between segregation and the timing of construction in a cox proportional hazard model

		City size restrictions			
	All cities	Bottom 50th percentile	50th to 80th percentile	80th to 95th percentile	95th and up
ln(total households)	2.935***	4.981***	3.625***	2.655***	1.755***
	(0.163)	(1.179)	(0.731)	(0.615)	(0.353)
Segregation	1.114	0.661	1.104	4.999***	34.435***
	(0.246)	(0.225)	(0.392)	(3.079)	(43.597)
Black pop. share	0.274***	0.113***	0.104***	0.082**	0.118
	(0.107)	(0.059)	(0.083)	(0.061)	(0.273)
Observations	1,188	608	351	178	51
Failures	823	279	318	175	51

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Coefficients reported in the table are hazard ratios. Each regressions includes census region fixed effects. Robust standard errors in parentheses. The median city in our sample has 253 households, while the 80th percentile had 995 households, and the 95th percentile had 4474 households.

hypothesized earlier. Unfortunately, our ability to examine racial gaps in access is limited by the fact that comprehensive household-level access data are not available for this time period.

Here, however, we can examine a few case studies by drawing on existing data from the Union Army Project and the Urban Transition Project. The Union Army project has generated shape files for the water and sanitation systems of Baltimore, Boston, Chicago, Philadelphia, Manhattan, and Brooklyn. These files include the year in which each main was installed, which allows us to generate a measure of exposure by assuming that households residing on the same street as the main had access to that main. Next we pair this with data from the Urban Transition Project, which has geocoded each dwelling in the 1880 census for 39 major cities.

Combine these two sources allows us to examine whether racial gaps in provision existed for these cities in 1880. Table 2 summarizes access to water and sewerage by race for each city. Interestingly, only two cities (Baltimore and Boston) fit the traditional narrative, where black households have less access than white households. For

Table 2: Access to water and sewerage by race and city, 1880

	Share with water access		Share with sewerage	
Baltimore	0.50	0.39	—	—
Boston	0.42	0.31	—	—
Brooklyn	0.77	0.82	0.20	0.27
Chicago	0.87	0.94	0.95	0.86
Manhattan	087	0.92	0.53	0.66
Philadelphia	0.78	0.80	0.04	0.04

each of the remaining cities, it actually appears that black households had weakly *greater* access to water and sewerage than white households. One reason for this could be that in these big cities, where rudimentary waterworks were constructed much earlier, the cycling of homeownership or other shocks that affected the spatial distribution of residents may resulted in a situation where black households had greater access.[9]

This highlights the importance of considering medium sized cities where initial provision is more closely related to 1880 demographics, since the census will fall closer to the initial construction date. On this front, Troesken's 2002 case study of sewers in Memphis, Tennessee is particularly informative. Memphis started construction of its sewers in the January of 1880. Troesken obtains a map of both the sewer system as of 1884 and the location of households as of 1880, which we reprint below. The areas of the city with large white populations are all connected, as are the handful of black households that live next to these white households. On the northern side of town, however, where there were very few white households the entire section of the city remained without sewer access until the 1890s.

---

[9] While there is a large modern literature on the spatial distribution of residents, these patterns were also present historically. For instance, Heblich *et al.* (2016) show that following the onset of industrialization the interaction between wind patterns and local pollution resulted in a shock to the spatial distribution of residents, such that wealthier individuals moved to the "upwind" areas of the city, which offered less pollution exposure.

Figure 6: Sewer access in Memphis, reproduced from Troesken (2002)

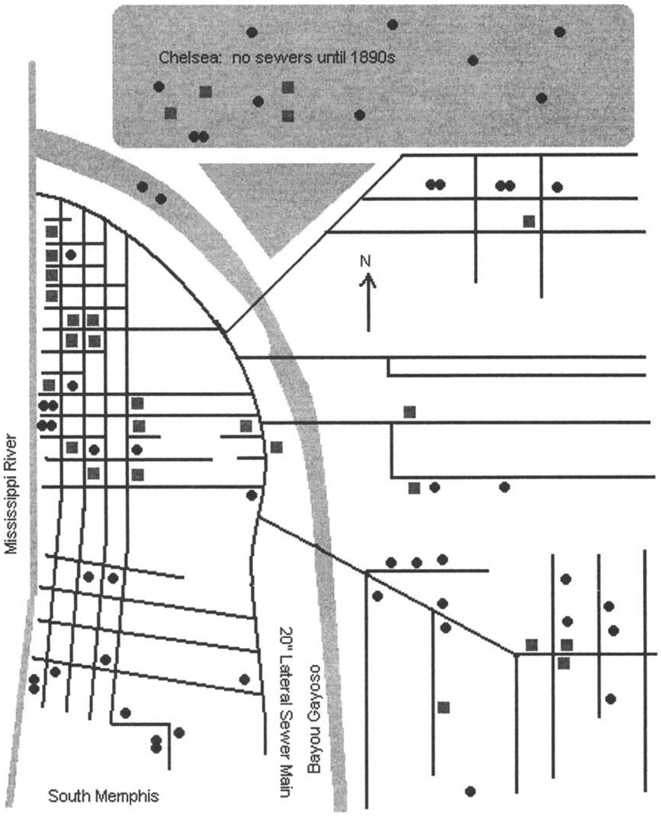


FIGURE 1
THE MEMPHIS SEWER SYSTEM, 1884

Key:
▬ - no sewerage services until the 1890s, including Chelsea and the area surrounding the Bayou Gayoso; ■ - white dwelling; ● - black dwelling; ●● - multi-race dwelling

Unfortunately, this is the limit of our ability to directly examine household gaps in access. For this reason, we now turn to our preferred approach of assessing proxies for access and how those proxies differ with segregation. This allows us to take a much more comprehensive look at our “exclusion” hypothesis.

4.3 Evidence from health outcomes

In this section, we set out to understand whether this potential lack of provision might have had negative health consequences. During the first half of the twentieth century, the adoption of water filtration technologies would ultimately generate large reductions in mortality (see Alsan & Goldin (Forthcoming), Ferrie & Troesken (2008)). In this sense, the initial wave of waterworks construction laid the necessary foundation for this public health movement. That is, once filtration technologies were developed, it would be possible to eliminate typhoid fever and other waterborne diseases by filtering the water that was already piped into homes and by continuing to increase household access to this clean water.

While much has been written about this mortality transition, which occurred between 1900 and 1950, it is important to note that the primary motivation of construction was not to purify the water. This is because effective water filtration technologies did not exist during the 1880s. The nation’s first filtration plant aimed at treating tainted water was the “Lawrence Experiment Station”, which was built in 1886 in Lawrence, MA (Alsan & Goldin, Forthcoming). Other filtration plants existed during this period, but they were aimed at combatting turbidity, discoloration, and bad taste rather than disease (Cutler & Miller, 2005). Anderson *et al.* (2018) provide additional support for this claim. In that paper, Anderson *et al.* trace out the history of clean water movements for 25 of the largest cities in the United States. Among these cities, Indianapolis and Providence were the first to invest in water filtration and those investments occurred in 1904. The median city invested in filters in 1913.

For chlorination, Jersey City was the first to chlorinate in 1908 and the median start for chlorination was 1913.

Baker's comments on the rapid expansion of waterworks during the 1890s are consistent with the interpretation that purification was not the primary motivation. As Baker writes:

> "Think of the effect upon the standard of living caused by the introduction of a public water supply! In place of the labor attendant upon lifting water by the old oaken bucket, the more prosaic hand pump, or of carrying water in pails from some spring or stream, only a turn of the faucet is now necessary in hundreds of communities to secure either hot or cold water on any floor of a dwelling. The labor saving thus secured, together with the increase in convenience, comfort and cleanliness, is too evident to need detailed mention, especially as both the old and the new are within the experiences of so many."

Largely due to the technological constraints of the time, initial water investment decisions were motivated by convenience rather than the control of disease. The systematic exclusion of blacks makes much more sense when the goal is to increase convenience rather than when the goal is to combat disease. When the goal is to combat disease, there is a stronger motivation to increase black access since humans can be carriers of waterborne diseases like typhoid fever. Thus, white households with access may remain vulnerable due to their interaction with unconnected black households. In the context of our model, local officials interested in combating typhoid fever and other waterborne diseases may value black household and white household access equally ($\alpha = 0.5$), which would undermine any empirical predictions about how segregation should affect the city's overall health. These are indeed the behaviors Troesken (2004) explores in the Jim Crow era of the early twentieth century as an explanation for why black health outcomes may have benefited substantially from public water and sewer investments.. When investment decisions are motivated by convenience, however, it is much more likely that local officials will undervalue black access. This is important because it suggests that $\alpha > 0.5$, and so we should expect the level of segregation in the city to generate variation in black vs. white household provision.

Although purifying the public water supply was not a primary motivating factor for construction, there are several ways in which the construction of a waterworks has the potential to improve public health. First, the newly built waterworks may have allowed the city to tap into water sources that were less likely to be tainted by improper sewage disposal, thus limiting exposure to waterborne diseases (see, for instance, Alsan & Goldin (Forthcoming) on water and sewerage access in Massachusetts). Second, residents with household connections to these waterworks no longer needed to collect and store water in pails. This lowered the cost of engaging in basic hygienic practices like hand washing. It also reduced the likelihood that stored water would become contaminated. Nevertheless, while these waterworks had the potential to improve public health, these improvements would be undermined when large segments of the population are excluded from accessing the waterworks. This is because piped water will offer very limited protection for any infectious disease that can be spread through human contact. Put another way, residents with water access will remain vulnerable to infection through their contact with residents without water access.

This generates an empirical prediction: if segregated cities were excluding blacks and thus under-investing in their infrastructure, then we should expect any health benefits associated with the construction of a waterworks to be decreasing as the level of segregation increases. To test this empirical prediction, we draw on complete count data from the 1900 census, as made available through IPUMS and Ancestry.com. In that census, each woman was asked about how many live births they have delivered and how many of those children are still living today. As in Logan & Parman (2018), we use this information to construct an indicator variable for whether an individual ever lost a child, which we interpret as a proxy for infant mortality. We rely on this proxy because comprehensive city-level data on black and white mortality by age and cause does not exist for this time period.

We analyze this relationship in a formal difference-in-differences framework. Our estimating equation is as follows:

$$Lost_{iac} = \alpha_0 + \beta_a + \gamma_c + \delta * Exposure_{iac} + \epsilon_{iac} \tag{11}$$

where $Lost_{iac}$ is an indicator that equals 1 if individual i, who is age a and resides in city c at the time of census enumeration has ever lost a child. The parameters β_a and γ_c are age and city fixed effects, respectively. The variable $Exposure_{iac}$ captures the share of fertile years (ages 18-45) that the individual spent in a city with water access. This is calculated by taking the individual's age at the time of enumeration and backing out how old they would have been at the time the waterworks was built. If the individual was under the age of 18 when the waterworks is built then they are assigned a 1. If the individual was 45 or older then they are assigned zero. The remaining individuals receive partial exposure corresponding to $\frac{45-age\ at\ time\ of\ construction}{45-20}$. In some specifications we will interact our exposure variable by the degree of segregation in the city. We run all of our analysis separately by race as given the dramatically different treatment of blacks during this period. Standard errors are clustered at the city level.

Identification of δ comes from the fact that different cities built their waterworks at different times and that, within a city, the construction of a waterworks generates differential exposure to water access based on the individual's age when the waterworks was built. One concern with this is that the waterworks may be built as part of a cluster of additional public health initiatives, all occurring at the same time. However, the actual timing of waterworks completion will be fairly decoupled from the timing of the whatever political process or change in public opinion initiated construction. It took substantial time between the initial decision to build a waterworks and the actual completion of that waterworks, with delays occurring for a variety of exogenous reasons.

Column 1 of Table 3 presents estimates of equation 11. The top panel presents results for white mothers while the bottom panel presents results for black mothers. Results from this specification indicate that investment in water decreased the likelihood of losing a child by roughly 4 percentage points for both black and white mothers. Note that our estimating equation includes city fixed effects, which absorbs the average effects of segregation, city size, and year built. In column 2 we interact our exposure variable with our segregation variable. In this specification, "Water Exposure" is interpreted as the effect of building a waterworks in an integrated city. For white mothers that resided in an integrated city, the effect of a waterworks construction is nearly identical to what we estimate in column 1. For the interaction term, we see that the health benefits are slightly larger for white mothers residing in segregated cities, though the coefficient is imprecisely estimated. In the bottom panel, however, we see large health effects for black mothers in integrated cities, and the interaction term is *positive* and statistically significant, indicating that the health benefits for black mothers residing in segregated cities were much lower than the effects in integrated cities. In Column 3 we run the analysis for the sample residing in cities above the median in terms of population, and in column 4 we restrict to the sample of mothers residing in the top 25% of the city size distribution and find similar effects.

4.4 Exploring the persistence of these health consequences

The main findings from the previous section are twofold. First, we showed that segregated cities were quicker to build their waterworks than integrated cities. Second, we showed that the health benefits of a waterworks were concentrated in integrated

Table 3: Difference-in-differences estimates of the relationship between waterworks construction and infant mortality

DV is whether mother has lost a child by 1900				
Sample:	All cities (1)	All cities (2)	Above med. city size (3)	Top 25 % city size (4)
Panel A: White Mothers				
Water Exposure	-0.037*** (0.012)	-0.033*** (0.012)	-0.026* (0.014)	-0.030 (0.018)
Water Exposure × Segregated City		-0.010 (0.008)	-0.020* (0.011)	-0.028 (0.017)
Sample Mean	0.375	0.375	0.374	0.375
Observations	1,704,294	1,704,294	1,543,628	1,336,236
R-squared	0.068	0.068	0.066	0.065
Panel B: Black Mothers				
Water Exposure	-0.041*** (0.012)	-0.126*** (0.022)	-0.149*** (0.031)	-0.132*** (0.044)
Water Exposure × Segregated City		0.093*** (0.021)	0.112*** (0.031)	0.088* (0.046)
Sample Mean	0.541	0.541	0.541	0.540
Observations	278,839	278,839	238,671	201,970
R-squared	0.068	0.068	0.062	0.058

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors (clustered at the city level) are reported in parentheses. Sample is restricted to black and white women between the ages of 18 and 55 who have had given birth to at least one child (at the time of 1900 census enumeration). Water exposure is the share of fertile years (ages 18 to 45) that the mother resided in a city with a constructed waterworks. Each regression includes city fixed effects, cohort fixed effects, and an indicator for whether the individual is white or not. Segregation is measured using the Logan-Parman segregation index.

cities. Together these findings are consistent with segregated cities building less extensive infrastructure because it was relatively less costly for them to exclude black households. However, as mentioned earlier, cities don't begin adopting clean water technologies en masse until after 1900. This lends itself to one more empirical test of the hypothesized exclusion mechanism that we think is underpinning our earlier results. Specifically, if cities underinvested in black neighborhoods because the perceived benefit of piped water was anchored more heavily on convenience than health, then this underinvestment should limit the city's ability to eliminate waterborne disease once those new technologies are developed. This is because once effective filtration technologies exist, it is still necessary to have a large share of the city's population connected to the filtered water supply to eliminate waterborne diseases. Thus, to capture the full benefits of filtration, cities will have to both invest in the technology itself and invest in extending their existing infrastructure. In segregated cities, if the pre-existing network of mains is less extensive, the necessary infrastructure investment will be relatively higher, which should in turn delay the city's ability to eliminate waterborne disease rates.

To examine this prediction, we draw on annual mortality data from annual mortality reports published by the U.S. Census Bureau. These reports, which tabulate important causes of death for registration states and cities, are available on an annual basis beginning in 1900.[10] The advantage of using these reports is that the underlying data conform to a common reporting standard. The disadvantage of using these reports is that the data are not comprehensive. Cities and states were only included in the published reports if the underlying data were deemed reliable. In 1900 there

---

[10] Registration states and cities are those with laws requiring that mortality statistics be collected. In contrast to England, which standardized and mandated the reporting of deaths in 1846, the United States left this decision to state and local governments. Several large cities and states passed mandatory reporting laws by 1900, and in that year the Census Bureau worked with those registration areas to establish uniform reporting standards. The result of this was the adoption of a standardized death certificate and the international classification standard, as well as the distribution of "The Manual of International Classification of Causes of Death", which cross referenced terms appearing in causes of death from 1890 and 1900 reports with the new uniform classification standard.

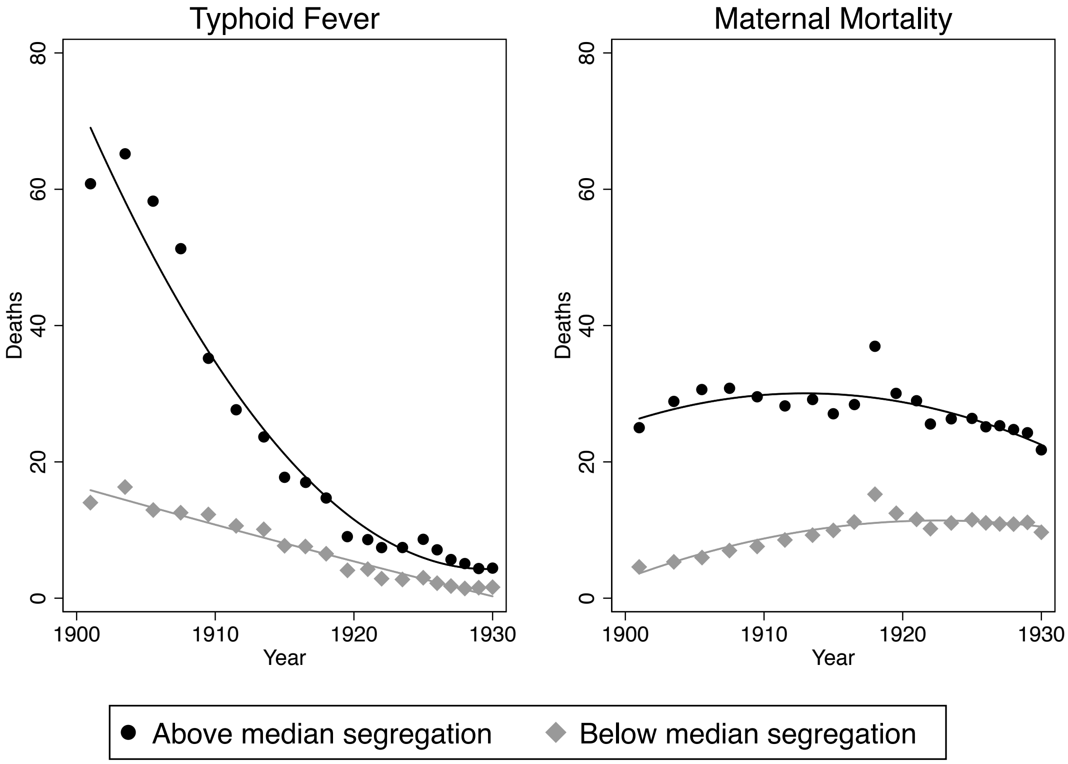
were 330 registration cities systematically collecting mortality data, but by 1920 the registration area included 662 cities spanning 41 states. We transcribe each report from 1900 through 1930.

Figure 7 examines the relationship between segregation and two measures of health between 1900 and 1930. The left panel plots typhoid fever deaths, a standard metric for assessing water quality and the prevalence of waterborne disease. There we see three important takeaways. First, we see that typhoid fever was nearly three times more prevalent in cities that were more segregated in 1880. It is worth noting that these mortality figures understate the significance of eliminating waterborne diseases. While typhoid fever had a relatively low case fatality rate, killing only 5 to 10% of infected individuals, contracting typhoid fever left an individual more susceptible to a host of other causes of death including, tuberculosis, heart failure, and kidney failure, and pneumonia and other respiratory infections. It is estimated that for each typhoid fever death that was prevented by water purification, an additional 4 to 7 deaths due to typhoid's sequela are also prevented (Cutler & Miller, 2005; Ferrie & Troesken, 2008). The second pattern we see in Figure 7 is that typhoid fever is not only more prominent in segregated cities in 1900, typhoid fever is more prominent in segregated cities throughout the entire study period. The final pattern we see is that by 1930 typhoid fever rates is largely eliminated in both segregated cities and integrated cities.

While these three patterns are consistent with what we would expect if segregation led to under provision of infrastructure, in turn hindering the city's ability to eliminate waterborne disease, it is perhaps useful to have a second point of comparison. For instance, one might be concerned that segregated cities were less developed in 1900 but between 1900 and 1930 they experienced relatively larger changes in income that could explain the patterns observed in Figure 7. One way to alleviate this concern is to examine non-waterborne causes of death to see if a similar pattern exists. As mentioned earlier, this is complicated by the fact that typhoid was directly related to many "non-waterborne" causes of death through its lingering sequela. One cause of death that is likely unaffected, however, is maternal mortality. Maternal mortality deaths largely resulted from hemorrhaging or infections picked up during the birthing process. To the extent that broader local economic development leads to improvements in the non-waterborne infectious disease environment or improved medical use, maternal mortality rates offer an interesting counterfactual that helps provider a broader context for the patterns observed with typhoid fever.

The right hand panel of Figure 7 plots the relationship between segregation and maternal mortality over time. While we do see that more segregated cities had more maternal mortality deaths in 1900, we also see little evidence of convergence between high and low segregated cities. We also see little evidence of a systematic decline in maternal mortality during this period. This tells us that the declines in typhoid fever, and perhaps more importantly the convergence, is unlikely to be due to systematic and broader economic changes occurring in segregated cities.

Figure 7: Binned scatter estimates of the relationship between segregation and city-level health



Notes: Mortality data come from annual reports. Sample includes 383 cities.

5 Conclusion

During the first half of the twentieth century, the United States experienced a dramatic decline in mortality as cities invested in clean water technologies. However, this public health movement would not have been possible without prior infrastructure investment, which connected households to a centralized water supply. This paper examines the extent to which the segregation of blacks influenced municipal investments, and in turn shaped the subsequent public health movement. We find evidence consistent with the narrative that segregated cities were quicker to build their waterworks and more likely to exclude black households. This exclusion, however, appears to be costly. We also find that segregated cities experienced much smaller declines in infant mortality, and segregated cities were much slower to eliminate waterborne diseases like typhoid fever. These results are consistent with segregation-induced exclusion undermining the city's ability to effectively control waterborne disease.

References

Alsan, Marcella, & Goldin, Claudia. Forthcoming. Watersheds in infant mortality: The role of effective water and sewerage infrastructure, 1880 to 1915. *Journal of Political Economy.*

Anderson, D Mark, Charles, Kerwin Kofi, & Rees, Daniel I. 2018. *Public Health Efforts and the Decline in Urban Mortality.* Tech. rept. National Bureau of Economic Research.

Antman, Francisca M. 2015. *For want of a cup: The rise of tea in England and the impact of water quality on economic development.* Tech. rept. Mimeo.

Baker, Moses Nelson. 1897. *The manual of American water-works.* Vol. 4. Engineering News.

Beach, Brian, & Hanlon, W. Walker. 2018. Coal smoke and mortality in an early industrial economy. *The Economic Journal*, **128**(615), 2652–2675.

Beach, Brian, Ferrie, Joseph, Saavedra, Martin, & Troesken, Werner. 2016. Typhoid fever, water quality, and human capital formation. *Journal of Economic History*, **76**(01), 41–75.

Cain, Louis, & Hong, Sok Chul. 2009. Survival in 19th century cities: The larger the city, the smaller your chances. *Explorations in Economic History*, **46**(4), 450–463.

Cain, Louis, & Rotella, Elyce. 2001. Death and spending: urban mortality and municipal expenditure on sanitation. *Pages 139–154 of: Annales de démographie historique*. Belin.

Chapman, Jonathan. 2018. The contribution of infrastructure investment to Britain's urban mortality decline, 1861–1900. *The Economic History Review*.

Clay, Karen, Lewis, Joshua, & Severnini, Edson. 2015. *Canary in a Coal Mine: Impact of Mid-20th Century Air Pollution Induced by Coal-Fired Power Generation on Infant Mortality and Property Values*. Tech. rept. Working paper.

Cutler, David, & Miller, Grant. 2005. The role of public health improvements in health advances: The twentieth-century United States. *Demography*, **42**(1), 1–22.

Ferrie, Joseph P, & Troesken, Werner. 2008. Water and Chicagos mortality transition, 1850–1925. *Explorations in Economic History*, **45**(1), 1–16.

Fogel, Robert W, & Costa, Dora L. 1997. A theory of technophysio evolution, with some implications for forecasting population, health care costs, and pension costs. *Demography*, **34**(1), 49–66.

Fogel, Robert William. 2004. *The escape from hunger and premature death, 1700-2100: Europe, America, and the Third World*. Vol. 38. Cambridge University Press.

Heblich, Stephan, Trew, Alex, & Yanos, Zylberberg. 2016. East side story: historical pollution and persistent neighborhood sorting.

Kesztenbaum, Lionel, & Rosenthal, Jean-Laurent. 2011. The health cost of living in a city: The case of France at the end of the 19th century. *Explorations in Economic History*, **48**(2), 207–225.

Kesztenbaum, Lionel, & Rosenthal, Jean-Laurent. 2017. Sewers diffusion and the decline of mortality: The case of Paris, 1880–1914. *Journal of Urban Economics*, **98**, 174–186.

King, Miriam L, & Magnuson, Diana L. 1995. Perspectives on historical US census undercounts. *Social Science History*, **19**(4), 455–466.

Logan, John R, Jindrich, Jason, Shin, Hyoungjin, & Zhang, Weiwei. 2011. Mapping America in 1880: The urban transition historical GIS project. *Historical Methods*, **44**(1), 49–60.

Logan, Trevon D, & Parman, John M. 2017. The national rise in residential segregation. *Journal of Economic History*, **77**(1), 127–170.

Logan, Trevon D, & Parman, John M. 2018. Segregation and mortality over time and space. *Social Science & Medicine*, **199**, 77–86.

McKeown, Thomas. 1976. The modern rise of population.

Preston, Samuel H, & Haines, Michael R. 1991. *Fatal years: Child mortality in late nineteenth-century America*. Vol. 1175. Princeton University Press.

Ruggles, Steven, Genadek, Katie, Goeken, Ronald, Grover, Josiah, & Sobek, Matthew. 2015. *Integrated Public Use Microdata Series: Version 6.0 [dataset]*. Minneapolis: University of Minnesota.

http://doi.org/10.18128/D010.V6.0.

Troesken, Werner. 2002. The limits of Jim Crow: race and the provision of water and sewerage services in American cities, 1880–1925. *Journal of Economic History*, **62**(03), 734–772.

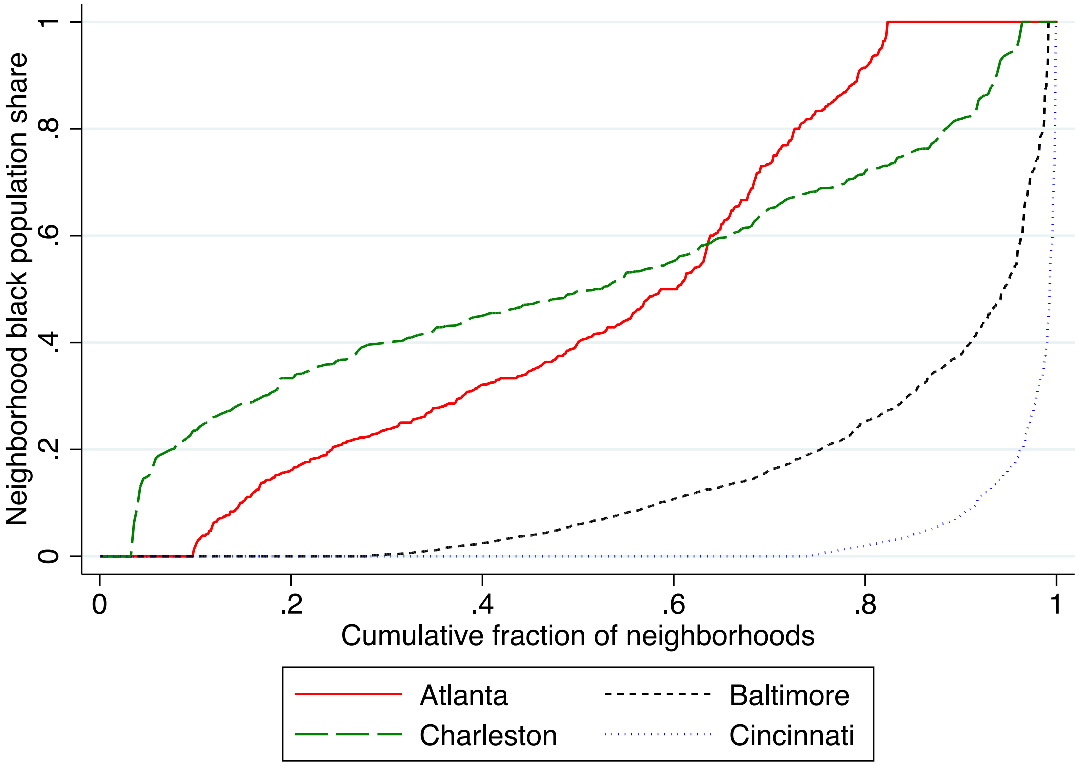
Troesken, Werner. 2004. *Water, race, and disease.* MIT Press.

A Empirical distributions of $g(.)$

In the theoretical model, the impact of segregation on water provision depends on the distribution of neighborhoods by racial composition, captured by the function $g(.)$. Several stylized versions of this function are provided in the main text. It is possible to construct approximations of $g(.)$ for several of the cities in our sample using fully geocoded versions of the 1880 complete count census created by John Logan (Logan *et al.* , 2011). These data contain geographic coordinates for every household in 39 cities matched to the census characteristics for the individuals in those households.

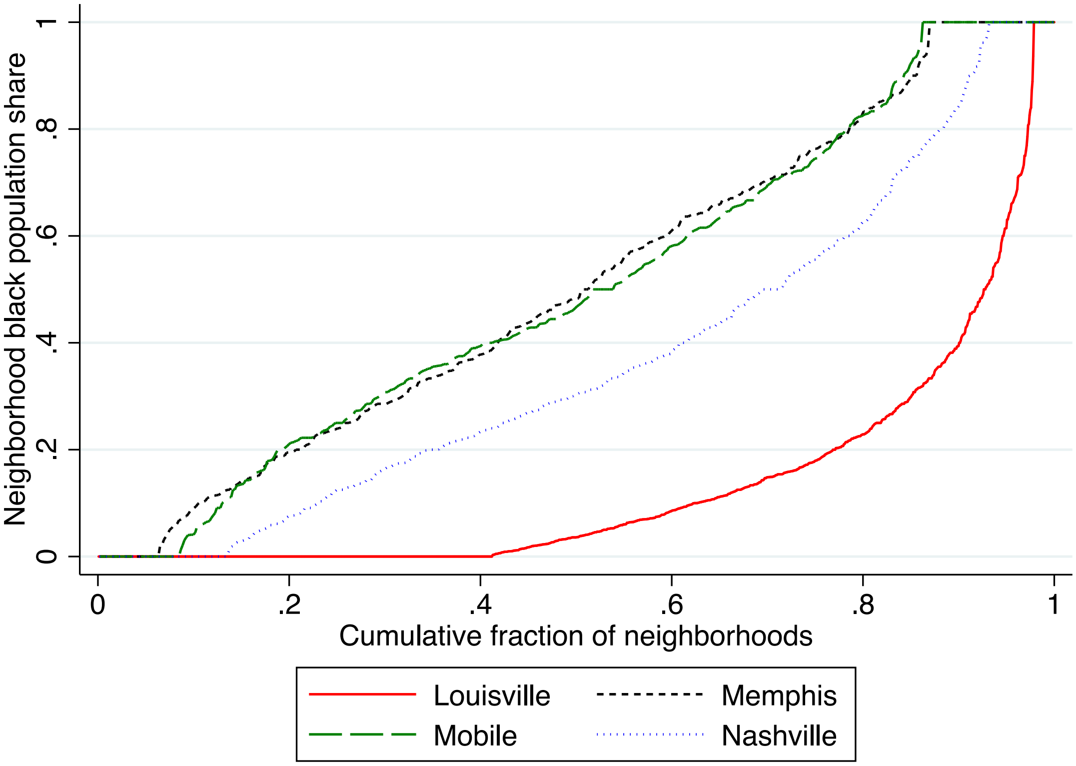
We create a tessellation of each city, a grid of squares with 150 meter sides covering the entire city. These squares are approximately the same geographic areas as the median census block. We then identify the races of all individuals living in each square to calculate the black population share of each square. Ordering these square 'neighborhoods' from lowest to highest black population share gives us an empirical approximation of $g(.)$. The resulting distributions for the cities in Logan's geocoded data containing sizable black population shares are provided in Figures 8, 9 and 10.

Figure 8: Empirical distributions of $g(.)$ for Atlanta, Baltimore, Charleston and Cincinnati in 1880



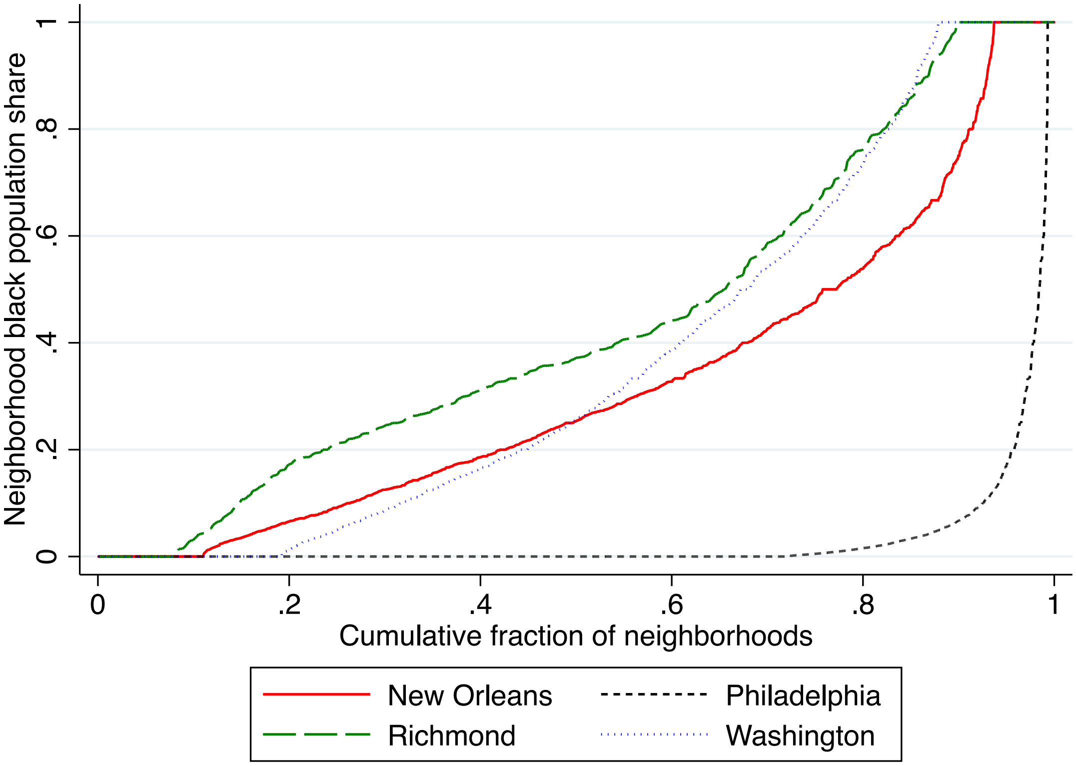
Notes: These data come from geocoded 1880 federal census data from Logan's *Urban Transition Historical GIS Project*. Neighborhoods are created with a tessellation of 150 meter by 150 meter squares covering the entire city.

Figure 9: Empirical distributions of $g(.)$ for Louisville, Memphis, Mobile and Nashville in 1880



Notes: These data come from geocoded 1880 federal census data from Logan's *Urban Transition Historical GIS Project*. Neighborhoods are created with a tessellation of 150 meter by 150 meter squares covering the entire city.

Figure 10: Empirical distributions of $g(.)$ for New Orleans, Philadelphia, Richmond and Washington, D.C. in 1880



Notes: These data come from geocoded 1880 federal census data from Logan's *Urban Transition Historical GIS Project*. Neighborhoods are created with a tessellation of 150 meter by 150 meter squares covering the entire city.

B Extending the model

B.1 Private provision without price discrimination

Suppose that the city is on a unit interval and the degree of segregation is described by the g function from the previous subsection. However, instead of public officials deciding water provision, suppose water is provided by a profit maximizing firm. The profit maximizing firm must decide what price to charge and how many miles of mains to build. Assume the firm cannot price discriminate, and must charge a single price for connecting to the water system. Further assume that all whites are willing to pay δ to connect to the water system, whereas all blacks are willing to pay θ, and that $\delta > \theta$. Since whites are willing to pay more, the firm will start construction of the mains in the whitest part of town, indexed by 0 on the unit interval.

The firm will either charge δ or θ to hook up to the water system. If the price were set below δ and above θ, then the firm could raise prices without losing any customers; the firm would never charge below θ because it is the minimum willingness to pay in the model. If the firm charges δ, then profits are

$$\Pi_\delta = \delta \int_0^m (1 - g(x))dx - pm - C, \tag{12}$$

where m is the miles of mains, p is the per unit cost of mains, C is the fixed cost of building the mains. The first order condition yields:

$$m^* = g^{-1}\left(1 - \frac{p}{\delta}\right). \tag{13}$$

This implies that miles of mains decrease as the price per main increases and that miles of mains increases as the willingness of whites to pay increases. Furthermore, mains increase with the share of whites. Segregation decreases water provision if the mark up is sufficiently low ($2p > \delta$). This is because if the mark up is low, then the private firm will stop constructing mains when it reaches the black neighborhood. If the mark up is sufficiently high, however, it will be worth constructing mains in the black neighborhoods to reach the few white customers that do exist. For this equilibrium, only whites connect to the water system.

The second possible price strategy is if the firm charges θ so that both black and white customers connect to the main. In this case,

$$\Pi_\theta = \theta m - pm - C \tag{14}$$

So long as the price per unit of main is smaller than θ, this will give us a corner solution of providing water to the whole city. In this equilibrium, the firm is already covering the whole city and segregation and the share of whites have no effect on main mileage.

Letting m^*_δ be the optimal miles of main under the δ pricing strategy, then the firm will pick the δ pricing strategy so long as:

$$\delta \int_0^{m^*_\delta} (1 - g(x))\, dx - \theta > pm^*_\delta - p. \tag{15}$$

Of course fixed cost must be sufficiently small so that profits are non-negative.

B.2 Private provision with price discrimination

Now suppose the firm can price discriminate, so that the firm charges all blacks θ and all whites δ so long as the firm builds a main to that particular neighborhood. Then profits become:

$$\Pi = \delta \int_0^{m} (1 - g(x))dx + \theta \int_0^{m} g(x) - pm - C \tag{16}$$

The first order conditions give us that:

$$m^* = g^{-1}\left(\frac{\delta - p}{\delta - \theta}\right) \tag{17}$$

Therefore, mileage of mains decreases with costs p. The comparative statics reveal that mileage increases with the WTP of whites if the costs per unit is greater than the WTP of blacks ($p > \theta$). Notice that if $p > \theta$, then whites are effectively subsidizing blacks. In the absence of whites in the neighborhood, building in the black neighborhood would not be profitable. An increase in the WTP of blacks increases main mileage so long as $\delta > p$, which is necessary for the water provision to be profitable. Water provision increases with the share of whites γ. An increase in segregation increases water provision so long as $\delta - \theta > 2(\delta - p)$ (or, alternatively, $2p > \delta - \theta$), which is to say if WTP of whites over blacks is at least twice as large as the mark up for whites. This result is similar to the role of segregation without price discrimination. In the case of no price discrimination, if the mark up for whites is high ($2p > \delta$), it is worth extending mains to neighborhoods with even a small number of whites. In the case of price discrimination, this logic still holds except except that the threshold is lower given that the firm receives revenues from the black households ($2p$ must be greater than $\delta - \theta$ rather than just δ).