Sewers and Urbanization in the Developing World*

Sean McCulloch,† Matthew Schaelling,‡
Matthew A. Turner,§ Toru Kitagawa¶

22 December 2025

Abstract: *We investigate the effects of sewer access on developing world cities. It is more difficult to move sewage uphill than downhill, so similar neighborhoods on opposite sides of drainage basin divides face different costs of sewer access. We identify the effect of sewers by comparing outcomes for neighborhoods on opposite sides of drainage basin divides. On average, sewers in Brazil, Colombia, South Africa, Jordan, and Tanzania cause large increases in population density and moderate changes in demographics. There is evidence for heterogeneous effects across countries. Estimates suggest that sewers are as important for the geography of cities as transportation infrastructure.*

JEL: O18, R3, L97, N11
Keywords: Sewers, Urbanization, Infrastructure

---

*We are grateful to Victoria Delbridge and IGC country staff in South Africa, Pakistan, Jordan, Ghana and Zambia for help with data collection. We are also grateful for helpful comments from Alex Rothenberg and from seminar participants at the University of Wisconsin, the European and North American meetings of the Urban Economics Association, and the meetings of the Latin American and China Urban Economics Associations. This research was supported by IGC grants BRA-23005 and XXX-23154 and by the Brown University PSTC, which receives funding from the NIH, for training support (T32 HD007338) and for general support (P2C HD041020). Any errors are our responsibility alone.

†Brown University, Department of Economics, Box B, Brown University, Providence, RI 02912. email: sean_mcculloch@brown.edu.

‡Brown University, Department of Economics, Box B, Brown University, Providence, RI 02912. email: matthew_schaelling@brown.edu.

§Brown University, Department of Economics, Box B, Brown University, Providence, RI 02912. email: matthew_turner@brown.edu. Also affiliated with PERC, IGC, NBER, PSTC, S4.

¶Brown University, Department of Economics, Box B, Brown University, Providence, RI 02912. email: toru_kitagawa@brown.edu.

1 Introduction

The economic logic of cities is simple. We are more productive if we work at higher densities than we prefer at our residences. Cities arise as the arrangement of work and residence locations that allow higher employment and lower residential density. It follows that our willingness to tolerate density is fundamental to how cities are organized. Sewer access has obvious implications for our willingness to tolerate population density, and so it is natural to suspect that it is important for the geography of cities.

We investigate the effects of sewer access on census tract population density and several other outcomes that describe tract level demographics and the built environment. We base our estimates on a quasi-experimental research design that derives from principles of wastewater engineering. Because it is more difficult to move sewage uphill than downhill, otherwise similar census tracts on opposite sides of drainage basin divides sometimes face different costs of sewer access. We use this intuition, and census tract level data, to estimate the effects of treating census tracts with better sewer access. We identify treatment effects by comparing rates of sewer access and outcomes for census tracts on opposite sides of drainage basin divides in Brazil, Colombia, South Africa, Jordan, and Tanzania.

It is not clear whether sewer access should increase or decrease population density. On the one hand, sewers eliminate the problems of on site management of sewage and free up land that would otherwise be used for outhouses, cesspools and cesspits. On the other hand, by making a place nicer, sewers may attract wealthier residents who demand more space. We estimate that increasing the number of sewer connections in an average census tract by one percentage point increases population density by about 6%. Using this estimate to evaluate small counterfactual sewer expansions suggests that expanding sewer networks has about equal, but opposite, effects on urban density as large expansions of transportation networks. This average effect appears to conceal cross country heterogeneity. There is suggestive evidence that sewer access *decreases* density in South Africa, and possibly Tanzania. Unsurprisingly, the higher population densities in sewered areas require bigger buildings. We estimate that increasing the number of sewer connections in an average census tract by one percentage point increases the number of households in an apartment by about 0.4

percentage points.

Looking at the effect of sewer access on tract demographics, we find that sewer access has no measurable effect on tract share literate, but that it has a modest effect on tract mean income. This effect is estimated imprecisely but suggests that universal sewer access in an unsewered tract about doubles tract mean income from about 933 to about 1930 USD2022 per month. On average, improving sewer access probably results in some displacement of relatively poor incumbents.

Sanitary sewers are impractical without access to piped water. There is also evidence that, absent garbage collection, garbage ends up in sanitary sewers and blocks them. Thus, we expect sewer access, piped water and garbage collection to often be provided jointly. This appears to be the case. We estimate that increasing the number of sewer connections in an average census tract by one percentage point increases the number of households with piped water by about 0.5 percentage points and the number with garbage collection by about 0.4 percentage points. Given higher baseline levels of piped water and garbage collection, this suggests that these services are provided jointly with sewers when they are not otherwise available. While increases to sewer access increase access to complementary services, they do not lead to blanket upgrading of local public services. In particular, electricity service does not vary with sewer access.

That sewer access causes changes to piped water and garbage collection raises the question of whether density responds to sewer access or to the joint provision of sewers, piped water and garbage collection. Restricting our estimation of the relationship between sewers and population density to a sample of tracts where piped water access or garbage collection is universal, or including piped water and garbage collection as control variables results in economically small changes in estimates of the effect size for sewers. Most of the effect on density is a result of sewer access, not piped water access or garbage collection.

According to the World Bank, about one third of the world's urban population did not have access to safely managed sanitation facilities in 2020, about the same proportion as live in slum conditions. Given the impact of safely managed sanitation on health and mortality, the need for improved sewer access is urgent, and improving such access is one of the United Nations' "Millennium

Development Goals." Yet, many cities also lack decent roads, sufficient public transit, adequate schools, and reliable electricity. Trade-offs between these services must be evaluated and made. Our finding that improved sewer access causes economically large increases in density should be of immediate use to policy makers evaluating such trade-offs.

Urban migration is among the best known ways to increase individual wages in developing countries (Gibson et al., 2014, Lagakos et al., 2020). Henderson and Turner (2020) estimate that for a typical resident of the developing world, moving to a location that is twice as dense increases household incomes by 32%. This leads us to ask why developing world countries are not urbanizing faster. One possibility is that developing world cities are difficult places to live, in part because they often lack basic sanitation. Our results strongly support this conclusion. By facilitating increased population density, sewer access can allow cities to accommodate more of the rural poor. Indeed, our estimates provide a foundation for the cost-benefit analysis of sewer expansions that includes benefits to rural migrants.

Despite its importance, the effect of sewer access on urban development has received little attention from researchers. There appear to be two reasons for this. First is the difficulty in organizing systematic descriptions of sewer networks. Sewers are underground, often old, and often administered locally, all factors that increase the difficulty of data collection. Second is the fact that sewers are not assigned to places at random, and the literature has failed to develop a quasi-experimental research design to address this problem that can be widely applied. We solve both problems. We exploit GIS technology to develop a quasi-experimental design using widely available census data and universally available digital elevation maps.

We provide two types of estimates. The first is a conventional TSLS/IV. At the census tract level, our treatment variable, share of households with sewer access, is continuous. This means that, outside of a homogeneous treatment effects framework, the TSLS/IV LATE involves an average of treatment effects that is not obviously of economic interest. To address this issue, we note that at the parcel level our treatment effect is binary. A parcel either has sewer access or not. We exploit this observation to estimate a parcel level MTE/LIV model with census tract level data by using a small variance approximation. Unlike the TSLS/IV estimand, the interpretation of the MTE/LIV based estimand is simple and economically meaningful. It is the census tract average of the parcel level effect of sewer access. In practice, the magnitude of the TSLS/IV and MTE/LIV estimates are about the same.

2 Literature

There is a large literature studying the effects of urban infrastructure. For example, Jedwab and Storeygard (2022) and Ghani et al. (2016) study the effects of highways and roads in India and Africa; Tsivanidis (2019) studies the effects of bus rapid transit in Bogota; Gendron-Carrier et al. (2022) studies the effects of subways all over the world; and finally, Allcott et al. (2016) and Lipscomb et al. (2013) study the effects of electrification in India and Brazil.

There is also a literature studying the effect of water quality on health outcomes, usually infant and child mortality, in the developing world (e.g., Ashraf et al. (2017), Galiani et al. (2005), Bhalotra et al. (2021)) and in the developed world during the industrial revolution (e.g., Anderson et al. (2018), Ferrie and Troesken (2008), Kesztenbaum and Rosenthal (2017), Ogasawara and Matsushita (2018)). These studies usually find large effects of improved water quality on health and mortality. A large public health literature also examines the effect of various interventions related to water, sanitation and hygiene. This literature generally finds large effects from the provision of water and sanitation (see Fewtrell et al. (2005) and Selendy (2011) for surveys).

Studies of sewers are rarer. Alsan and Goldin (2019) study late 19th century Boston and find a large reductions in infant mortality from the joint roll-out of municipal water and sewer systems, but no evidence that people sorted into places with better water and sewer service on the basis of observable demographics. Anderson et al. (2018) examine the effect of sewer system construction in 25 US cities in the early 1900s and, contrary to Alsan and Goldin (2019), find no relationship between measures of mortality and sewage treatment or the interaction of sewage treatment and water treatment.

To our knowledge, Gamper-Rabindran et al. (2010) is the only paper to explicitly study urban sewer systems in the developing world. This paper considers a municipality-year panel of Brazilian data reporting infant mortality and municipal level measures of water and sewer access. They find that access to piped water, but not to sewers, has a large effect on infant mortality. Only Coury et al. (2022) explicitly considers the relationship between sewer construction and urban development. Coury et al. (2022) investigates the effect of expansions of the Chicago water and sewer network in the late 19th century on the price of residential land. They find that sewer and water access more than doubles land prices.

Summing up, the available evidence suggests that sewer access has beneficial effects on cities and neighborhoods. The evidence for more specific effects is thin and based on 19th century US cities. Coury et al. (2022) find that sewer access leads to large increases in land prices in late 19th century Chicago. It is natural to suspect that these prices increases were associated with an increase in density. Extrapolating from the Alsan and Goldin (2019) finding that people do not sort on the basis of local variation in municipal water and sewer service suggests that people should not sort on the basis of sewer access. However, there is little evidence about the magnitudes of these effects in contemporary developing world cities, or about heterogeneity in these effects across places. These are the questions we address.

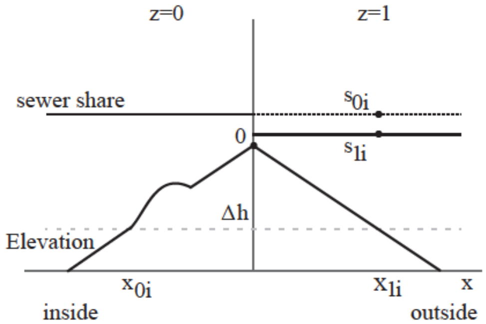
3 Identification

The movement of wastewater in sewers is sensitive to variation in elevation that is irrelevant for most other activities. Gravity sewers require a grade of about 1:200 (1 unit of drop per 200 of horizontal). While the details of pipe size, shape, and interior smoothness can partially compensate for vertical drop, in general, at a grade of about 1:200 solids settle out of the flow and block the pipe (Mara, 1996). For reference, athletes will generally perceive a playing field as sloped only once it has a grade of more than 1:70 (Aldous, 1999).

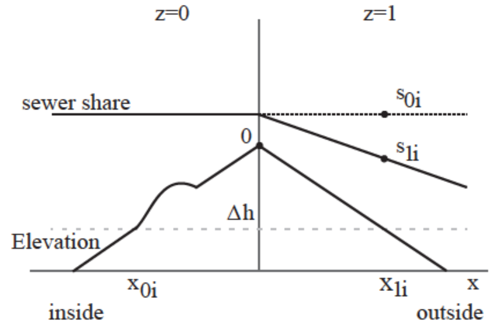
These two facts motivate the identification strategy illustrated in figure 1. The peaked dark line in this figure describes the elevation profile along an axis horizontal to a drainage basin divide at $x = 0$. The region to the left of $x = 0$ is "inside" the central city drainage basin and drains downhill to the sewer system servicing the CBD. The region to the right of $x = 0$ is "outside" and cannot reach the central city sewer network without traveling uphill.

Moving sewage across a basin divide is difficult and may be accomplished in three ways (Mara, 1996). First, by burying sewer pipes more deeply. Residential

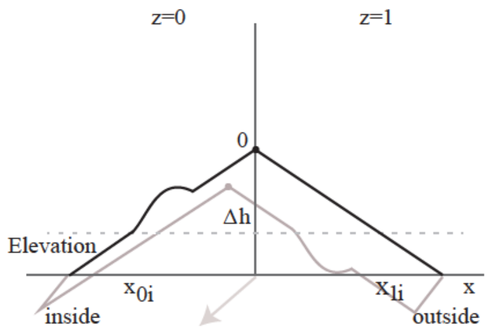
Figure 1: Identifying treatment effects around a stylized basin boundary



(a)



(b)



(c)

Note: *Elevation and sewer share profile in the neighborhood of a drainage basin divide. The basin divide is at the top of the hill, at $x = 0$. Displacement left is "inside" and towards the nearest established sewer system. Displacement right is "outside" and wastewater in this region must travel uphill to reach the nearest sewer network. (a) Crossing the basin divide is a discrete shock to the cost of sewer access. (b) Crossing the divide increases the cost of sewer access continuously with distance to divide. (c) Illustration of variation in elevation independent of x.*

sewer lines are typically buried between two and eight feet deep, while mains are sometimes hundreds of feet deep. Recalling that a sewer needs a grade of 1:200, burying a sewer 10 feet deeper allows an extra 2000 feet of horizontal travel in flat terrain. Second, if the topography allows, following an indirect route approximately along an elevation contour to the inside of the basin allows the substitution of downhill, horizontal travel for climbing. Third, it is also possible to build pumping facilities to lift the sewage over the basin divide. There is also the possibility of building a new sewer network to serve the relevant drainage basin and thus avoid moving sewage up and across a basin divide altogether. All four possibilities are costly. Crossing a drainage basin divide from a basin with sewer service to one without increases the cost of sewer access.

Summing up, for places on the outside of a drainage basin divide, the cost of reaching the central city sewer network will generally increase rapidly with the horizontal and vertical distance that sewage must cover to reach the basin divide. Conversely, for places on the inside of the basin divide, horizontal and vertical distance from the divide should have less impact on the cost of sewer access, or none at all.

As we will see, drainage basin divides are usually almost unnoticeable landscape features. From this it follows that locations close to, but on opposite sides of a drainage basin divide should be similar in their suitability for urban use, except that sewers will be more costly for outside locations. This suggests that for locations close to a drainage basin divide, being inside or outside the basin is a source of quasi-random variation in the cost of sewers. Our research design is organized around comparing census tract level sewer access and demographics in nearby tracts on opposite sides of a drainage basin divide.

The three panels of figure 1 inform the exercise of translating this intuition into an econometric specification. The horizontal axis, x, is displacement along a horizontal axis perpendicular to a drainage basin divide, henceforth simply "horizontal distance." In all four panels, locations to the left of zero are inside and are uphill from the sewer system serving the central city. Locations to the right of zero are outside, and sewage in these locations must travel horizontally and vertically to the basin divide before it can drain to the center.

Define a binary variable $\mathbb{1}(\text{Outside})(x)$ equal to one for x outside, and zero otherwise. Let s indicate the share of households in a location with sewer access, and define $\Delta h(x) \geq 0$ as meters of descent required to reach x from the top of the basin divide at $x = 0$. Thus, $\mathbb{1}(\text{Outside})(x)x$ and $\mathbb{1}(\text{Outside})(x)\Delta h(x)$ are the horizontal and vertical distance required to reach the inside of the central drainage basin from location x. We consider $\mathbb{1}(\text{Outside})(x)$, $\mathbb{1}(\text{Outside})(x)x$, and $\mathbb{1}(\text{Outside})(x)\Delta h(x)$ as instruments.

Our measure of sewer access is the share of households in a census tract reporting that they have access to a public sewer. The size of census tracts varies by country, but they are on average about one quarter square kilometer. At this scale, it is possible that the cost shock to sewer construction will appear instantaneous when we cross the basin divide. This case is illustrated in panel (a). In this figure, we suppose that sewer share, s, does not depend on x, except at the basin divide, where the cost of sewer access increases, and the share of houses reporting access to a sewer declines as a step function.

We would like to estimate how an outcome y depends on sewer access. If panel (a) is an accurate description of the world, then we can do this by comparing the size of the step down in sewer access at $x = 0$ to the corresponding change in y.

It is hard to have a strong prior about the spatial scale over which the basin divide cost shock will operate, and there are reasons to think that it will not operate as sharply as illustrated in panel (a). For example, the area near a drainage basin divide is often quite flat. In this case, a few hundred horizontal feet outside of the basin divide may involve a drop that can be accommodated by burying sewer lines a little deeper. In such cases, we expect sewer access to decline smoothly with distance to the basin divide.

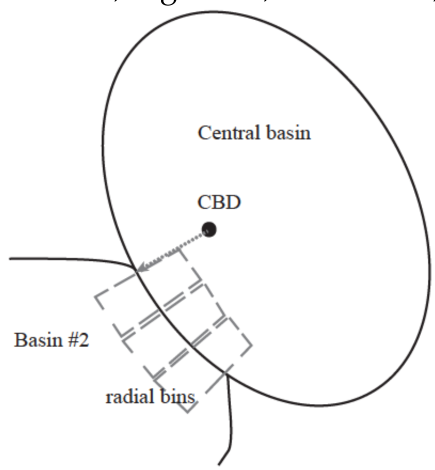
The case when sewer access declines continuously as we move further outside the central basin is illustrated in figure 1(b). The intuition behind panels (a) and (b) is similar, but the implied econometric model is not. Panel (a) suggests a discrete treatment while in panel (b) treatment is continuous, and the econometrics of estimating treatment effects with continuous and binary treatments are quite different.

Figure 1(c) illustrates a final point about our identification strategy. In reality, and unlike what we illustrate in the top two panels, our data will lie on a strip rather than a line. This means that there will be variation in elevation, holding distance to the basin divide constant. Therefore, we can estimate the effect of elevation on sewer access, conditional on x. Holding x constant, we expect vertical distance to have a larger effect on sewer access outside the basin divide than inside.

4 Central Basins

We now turn to the problem of defining an empirical analog to the illustrations in figure 1. Figure 2 illustrates our approach. The central ellipse in this figure describes the drainage basin containing a central city with a sewer system. This is the "central basin." All points in this basin drain to the same point, and so, in principle, can be served by the same sewer network. Any point

Figure 2: Illustration of basins, segments, radial-bins, and “inside” indicator



Note: *The central ellipse describes the drainage basin containing a center city. The boundary of this drainage basin is the central basin divide. A location is "inside" or "outside" as it lies inside or outside the central basin. The central basin generally abuts other drainage basins. The portion of the central basin divide which divides a particular pair of basins is a "segment" of the basin divide. We divide the area near the basin divide into "radial-bins"(sometimes "π-bins"). To construct these bins, we divide the central basin divide into two kilometer long intervals, starting from the point on the basin divide nearest the city center. A radial-bin is the area within 2KM of one such interval.*

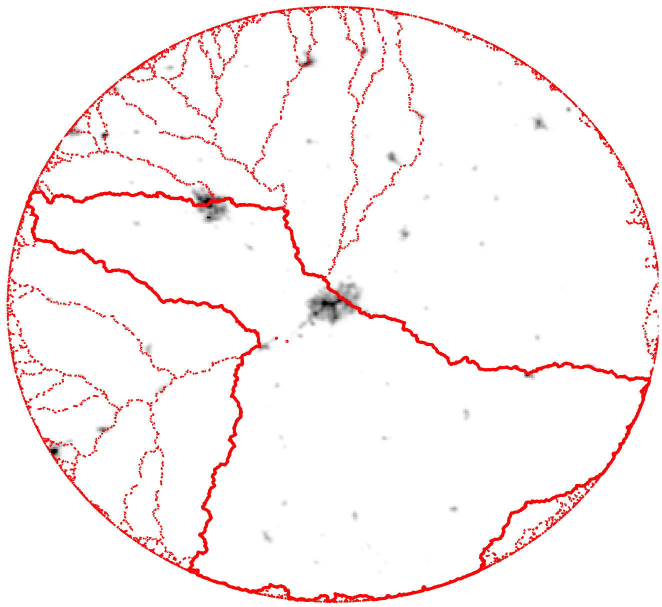
not in the central basin, by definition, does not drain to this point. The boundary of the central basin is the “central basin divide” and a location is “inside” or “outside” as it lies inside or outside the central basin.

The UN DESA World Urbanization Prospects data report the coordinates of the centers of all cities that have a population of 300,000 or above in 2018 (UN DESA Population Division, 2018). We restrict attention to cities in the UN Data for countries where we have census data and maps: Brazil, South Africa, Tanzania, Jordan, and Colombia. These are medium to large cities and all central cities in our sample have at least some sewer service near their centers.

We next download the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) (NASA et al. (2022)) digital elevation map. These data report the elevation of most of the Earth’s surface at a spatial resolution of about 30m by 30m.

From the ASTER data, we clip out a circle of radius 75KM centered on the CBD of each sample city. Each such circle is an elevation map of one of our cities and its hinterlands. This done, we draw all drainage basins within a 75KM radius of

Figure 3: Drainage basins containing Cascavel, Brazil



Note: *Dashed Red lines indicate drainage basins boundaries based on the* ASTER *digital elevation map. The solid red line indicates the basin boundary for the basin containing Cascavel, Brazil.* VIIRS *lights at night shows city extent. The disk has a radius of* 75KM.

the center of each city using an ArcGIS utility. Finally, we identify the drainage basin containing the center of each city. These are the central basins, and their boundaries are the central basin divides.

We rely on census data at the tract level to describe sewer access and outcome variables. Our unit of observation is a census tract, and our sample includes only tracts close to a central basin divide. We define a census tract as inside or outside depending on whether its area weighted centroid is on the same side of the closest central basin divide as the central business district.[1] Because all central basins contain a sewer network, this definition guarantees that an inside tract can drain to a central city sewer network.

Figure 3 is an empirical analog of figure 2 and illustrates basin boundaries around Cascavel, Brazil. Solid red shows the boundary of the central basin and red dashed lines indicate the boundaries of other drainage basins. Shading is

[1]We experimented with assigning census tracts to be inside or outside on the basis of their lights based centroid. This did not change our results.

based on lights at night and shows the scale of the city relative to the various basins.

In figures 2 and 3, each central basin abuts other drainage basins. The portion of the central basin divide which divides a particular pair of basins is a "segment" of the basin divide. For econometric purposes discussed later, we divide each central basin divide into these segments. We also divide the area near the basin divide into "radial-bins," which we sometimes abbreviate to "π-bins." To construct these bins, we divide the central basin divide into two kilometer long intervals, starting from the point on the basin divide nearest the city center. A radial-bin consists of all census tracts with centroids within 2km of one such 2km interval. We define "segment-bins" analogously on the basis of basin segments rather than two kilometer intervals.

Because two central basins may be adjacent, in theory, our notion of inside and outside can be undefined. However, it is rarely ambiguous in practice. For tracts for which the closest basin divide segment divides two central basins, on average the more remote cbd is three times as far away as the closer one. Therefore, for tracts for which the closest basin divide separates two central basins, we define inside and outside on the basis of the closest of the two city centers.[2] [3]

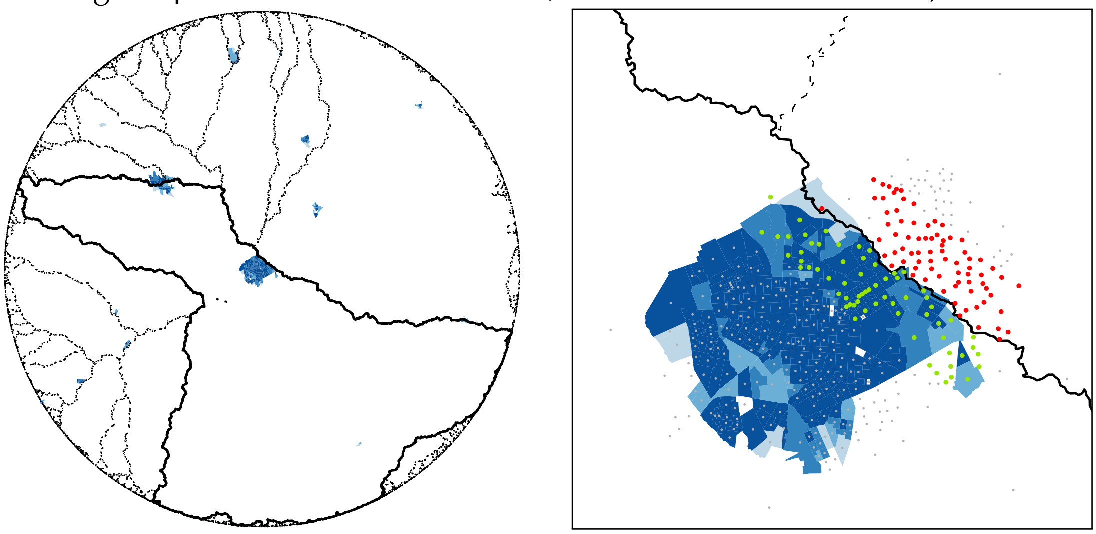
Looking carefully at figure 3, we see that our basin drawing algorithm sometimes constructs incoherent basins at the edge of the map disc. For this reason, we exclude from our study the region within 6km of the edge of these discs, or conversely, more than 69km from the center of the city.

The drainage basins that contain the coordinate of the cbd for our cities are sometimes too small to contain a meaningful share of the city's population

[2]An alternative approach to this problem is to drop all basin divides that divide two adjacent central basins. This reduces our sample size somewhat, but does not qualitatively change our results.

[3]Another problem may arise arise when small towns lie close to, but outside a central basin divide, and are far from the main cbd. If these small towns have sewer networks, then for census tracts close to these small towns, being inside the central basin probably places them farther from the nearest sewer network. We experimented with alternative definitions of inside that address this problem. These alternatives face the following problem. We can only measure sewer access with the same census data that we use to define our treatment. Thus, picking out small, highly sewered towns, near the basin divide relies on the same data we use to construct our treatment variable. Therefore, any definition of inside based on these data implicitly requires that we condition on an endogenous variable. Given this, we do not pursue alternative definitions of inside and outside.

Figure 4: Sewer share and "inside/outside" around Cascavel, Brazil



Note: *Full basin and close up of Cascavel, Brazil. Darker blue indicates a larger share of households reporting a toilet connected to a public sewer. Dots indicate census tract centroids. Centroids for which the closest basin divide is not the central basin are excluded (light gray), as are centroids that are more than 2KM from the central basin divide. "Inside" centroids are green, and "outside" centroids are red. Because basin boundaries are often incoherent near the edge of the 75KM disk of the DEM we work with, tracts with centroids more than 69KM from the city center are also excluded from the sample.*

(recall the UN DESA data reports on cities with a population above 300,000). To resolve this problem, we define our central basins as the union of drainage basins that intersect a disk of 2KM radius centered on the CBD.[4] This guarantees that no point on the central basin divide is closer to the CBD than 2KM.

5 Sewers and outcomes

The Brazilian census asks households if they have a toilet, and whether this toilet drains to a sewer, septic tank, a ditch, a pit, or surface water. In this way, the census provides an indicator of "connected to a sewer." These data are publicly available, aggregated to the "sector" (about the same size as a US census block group). Equivalent questions appear in the census forms of

[4]To see why this creates a problem consider two central basin segments, one about 100 meters from the CBD, and one 10KM from the CBD. For the first, displacement inside the basin divide is displacement towards the CBD for about the first 100 meters, and then it is displacement beyond and away from the CBD. In the second case, displacement away from and inside the basin divide is towards the CBD for about 10KM. Pooling these two types of basin divides complicates the interpretation of horizontal distance.

Colombia, Tanzania, Jordan, and South Africa, and are available at about the same spatial resolution. Appendix B provides more detail. These questions allow us to calculate the share of households in a census tract with access to a sewer for our entire sample of tracts. This is our treatment variable. We would like to know the effects of changes in the share of tract households with sewer access.

To calculate the vertical distance between each tract and the central basin divide, we calculate the height of the basin divide for each tract as the highest elevation of any tract centroid in the same radial-bin as the target tract. We then calculate the vertical rise required to reach the basin divide, $\Delta h(x)$, as the elevation difference between the target tract centroid and this radial-bin maximum.

Table A1 provides a country-by-country breakout of our data. Brazil accounts for the largest share of cities in our data, but the number of Colombian and Brazilian tracts in our data is about the same. South Africa has about the same number of cities as Colombia, but about a quarter the tracts. The samples for Jordan and Tanzania are smaller. The average over tracts of sewer share is about 0.7 for Brazil and Columbia, 0.8 for South Africa, 0.4 for Jordan, and about 0.08 for Tanzania.

Having defined our treatment as "share of tract households with sewer access," it is of interest to understand the alternatives to sewer access. Brazil, Jordan, South Africa, and Tanzania (but not Colombia) provide more detail about unsewered households and table A19 organizes these data. The two main alternatives to sewer access are cesspits and septic tanks, respectively a hole in the ground (possibly lined) and a lined tank in the ground. A third category consolidates "other" and "none." For the countries where we can refine the "not sewered" category, our estimation sample describes about 5.9m households with an average of 3.3 people per household. Of these, about 67% have sewer access and about 28% have a cesspit or septic tank. The remainder have no sanitation facilities or some other arrangement.

Septic systems are common in low density development in the US. These systems consist of a septic tank and a large, highly regulated drain field. The high population densities observed in our census tracts means that the septic tanks that prevail in our sample must be quite different from those in the rural US. Sanitation for most of the unsewered households in our sample is probably primitive.

Figure 4 is a heat map illustrating the incidence of sewer access for the Brazilian city of Cascavel. Polygons describe the extent of census tracts, with darker blue indicating a larger share of households reporting sewer access. Basin boundaries are black lines. Dots indicate census centroids. Census centroids are green if inside the central basin, red if outside, and gray if excluded from our sample. The right panel is a close-up of the area around the central city. It shows that sewer access drops dramatically across the central basin divide.

The Brazilian, Colombian, South African and Tanzanian censuses report population by tract. Because we have GIS maps of tract boundaries, we can also calculate tract area, and hence tract population density. These four censuses also report responses to questions about availability of piped water, garbage collection, electricity, and whether the household lives in an apartment. These censuses also report information about educational attainment or literacy that we use to create a standardized measure of the share literate. Because Brazil only reports on literacy, more detailed measures of educational attainment are not possible for the full sample. Only Brazil and South Africa report on household income, so we restrict our analysis of income to these two countries. Jordan's census reports only the count of households, so for Jordan we use household density in place of population density. Appendix B describes our census data in more detail.

Summing up, our data describes census tract equivalent units in Brazil, Colombia, South Africa, Tanzania and Jordan. We restrict attention to tracts whose centroids are (1) within 2km of the nearest central basin divide, (2) within 69km of the city center. We assign sewer share, population density and other outcomes to tracts using census data.

Table 1 describes our data. Column 1 reports on all tracts with centroids that fall in the central basin of one of the 95 cities in our sample. This column describes the cities of interest. This sample consists of about 246,000 tracts and 4,200 radial-bins. Column 2 describes all tracts that lie within 2km, inside or outside, of the central basin divide for one of these 95 cities. These are the tracts we will use to estimate the effects of sewer access. This sample consists of about 51,000 tracts and about 1,900 radial-bins. We refer to the sample described in column 2 as the "estimation sample."

As expected, the estimation sample is about one KM from the basin divide on average and is about evenly split between inside and outside tracts. The tract mean sewer share is about 0.68. The corresponding shares for piped water and garbage collection are 0.89 and 0.92, so these services are more widely available than sewers. For later reference, 53% and 64% of tracts in our sample have 100% coverage of piped water and garbage collection.

On average, a tract in the estimation sample is about 12KM from the CBD. By construction, tracts in the estimation sample are further from the CBD than those in the cities sample. In spite of this, mean distance to the CBD is larger in the cities sample. The table presents tract weighted averages, and cities located in large drainage basins have more tracts that are further away. As a result, restricting to tracts within 2KM of the basin divide drops many more tracts in large basin cities than in small basin cities, reducing the average distance to CBD when pooled across cities.[5]

Dividing total population in the estimating sample by total area, we have an average population density of about 1940 per square kilometer. These are moderately high urban densities. Tract mean population densities are much higher. Because dense areas contain many small tracts, dense areas are overweighted in these tract averages, and so tract weight density is higher.[6] The next two rows of the table report income per month and share literate. As described above, the income data reflects only Brazil and South Africa, and the share literate excludes tracts in Jordan.

[5]Within each city, the sample restriction does increase the average distance to the CBD as anticipated.

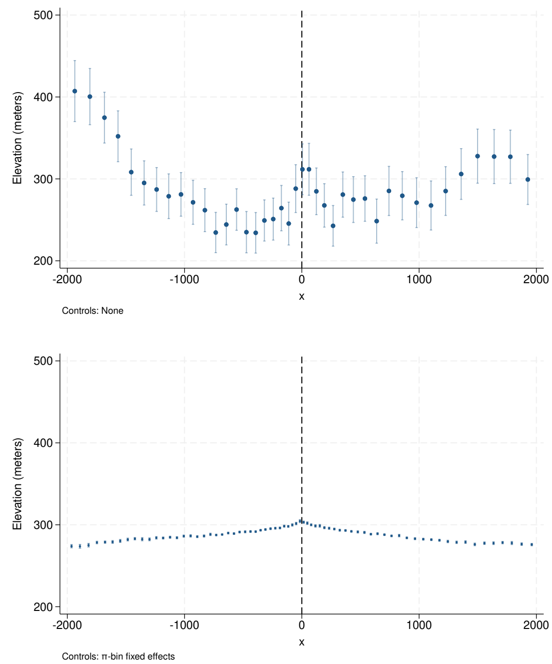
[6]To see the importance of this, consider a city with just two tracts: the first tract has 250 people and geographic area $= 0.01$ KM^2, and the second tract has 250 people and area 2KM^2. Then the average population density for the whole city is $(250 + 250)/(2 + 0.1) \approx 249$ people per KM^2 but the average of the tract level densities is more than 12,500 people per KM^2 ($\frac{1}{2}(\frac{250}{0.01} + \frac{250}{2}) \approx 12{,}563$).

Table 1: Descriptive statistics

	(1) CBD Basin + 2KM	(2) ± 2KM Basin Divide
Num cities	95	95
Mean area cbd basin (KM^2)	1,374	1,374
Num segments	844	559
Num π-bins	4,187	1,896
Num tracts	246,026	50,805
Share inside	0.90	0.54
	(0.30)	(0.50)
Mean tract area (KM^2)	0.63	0.22
	(5.79)	(1.77)
Mean dist to CBD (KM)	13.62	12.37
	(15.22)	(16.34)
Mean log dist to CBD (m)	8.95	8.74
	(1.11)	(1.16)
Mean dist to basin divide (KM)	11.48	0.86
	(9.33)	(0.59)
Sewer share	0.75	0.68
	(0.33)	(0.37)
Mean num people in a tract	333	428
	(365)	(393)
Pop density (persons/KM^2)	28,395	22,031
	(34,635)	(27,012)
Income (per month, 2022 USD)	968	933
	(921)	(832)
Share literate	0.94	0.92
	(0.29)	(0.40)
Mean share piped water	0.92	0.89
	(0.22)	(0.25)
Share with 100% piped water	0.67	0.53
	(0.47)	(0.50)
Mean share trash collection	0.92	0.92
	(0.23)	(0.23)
Share with 100% trash collection	0.71	0.64
	(0.46)	(0.48)
Share electricity	0.98	0.97
	(0.11)	(0.11)
Share apartment	0.27	0.27
	(0.31)	(0.31)
Elevation (m)	939.12	296.26
	(791.46)	(527.53)

Note: *Column 1 describes all tracts with centroids in one of the 95 central basins that make up our sample, or less than 2KM outside. All cities lie in Brazil, Colombia, Jordan, South Africa, or Tanzania. Column 2 describes our main estimation sample. It consists of all tracts that fall within 2KM of the central basin divide for sample cities. Income data based only on Brazil and South Africa. Literacy data excludes Jordan.*

Figure 5: Mean tract centroid elevation and conditional elevation as a function of distance to the nearest basin divide



Note: *Mean elevation by distance to basin divide at $x = 0$; raw data (top), and net of radial-bin means (bottom). Figures are based on the estimation sample described in column 2 of table 1. On average a central basin divide is not a dramatic feature of the geography.*

6 Descriptive results

Figure 5 shows empirical analogs of the elevation profile in figure 1. Each panel presents a binscatter plot of the mean bin elevation as a function of the distance from each tract centroid to the nearest point on the central basin divide. As in figure 1, the basin divide is at $x = 0$, with left of zero inside the central basin and right outside. Each panel is based on the estimation sample described in column 2 of table 1. The top panel shows unconditional bin means. In this figure, the expected high point at the basin divide is on average lower than interior tracts. In the bottom panel, we repeat the exercise but with elevation net of the radial-bin mean. We now see the expected peak at the basin divide clearly.

This figure demonstrates that the drainage basin divides are not dramatic geological features in two senses. First, comparing the bottom panel of figure 5 to the panel above, we see that the variation in elevation associated with a few KM of horizontal travel around a basin divide is small relative to total variation in elevation. Second, the basin divides are small features in an absolute sense. On average, traveling 2KM from a basin divide, whether inside or outside, involves a descent of about 30m. Thus, the average grade along a 2KM path extending from 2KM inside the basin divide to 2KM outside is about 1:70, the grade at which athletes begin to notice a playing field is sloped.

The peak at $x = 0$ visible in the bottom panel of figure 5 is by construction. Basin divides lie at local high points. That we cannot see the basin divide in the top panel indicates that, unless we restrict attention to tracts in the same radial-bin, we are not exploiting variation in elevation across tracts as we intend. This motivates our reliance on within radial-bin variation in our estimations.

Whiskers in the bottom panel of figure 5 describe variation in elevation around the mean, conditional on distance from the basin divide. This is variation in elevation holding horizontal distance constant. These tight confidence intervals indicate that, not only are basin divides almost unnoticeable features of the landscape on average, but that there is also not much variation around this average. There are not many basin divides where elevation changes dramatically in a neighborhood of the divide.

Figure A1 shows the correlation between the share of households in a tract with sewer access and the logarithm of tract population density. The figure is a binscatter plot, and so the slope reflects means in the data. Throughout most of the range of sewer access, the relationship is approximately linear, and the slope indicates an elasticity around one. That is, each 1 percentage point increase in sewer access is associated with a 1% increase in population density.

Figure 6 illustrates the empirical variation relevant to our first identification strategy. All three panels are binscatter plots reporting means for a different variable as a function of distance to a central basin divide, net of radial-bin means.

For reference, the top panel repeats the bottom panel of figure 5, but with the y-axis rescaled. The middle panel shows changes in sewer share as a function of distance to the central basin divide, net of radial basin means. We see a trend break in sewer share at $x = 0$, and possibly a small step down. This is consistent with our intuition about the costs of sewer construction and with the cross-divide decrease in sewer access in the right panel of figure 4. Thus, the middle panel of figure 6 confirms that crossing a basin divide decreases the prevalence of sewers. However, the functional form of this relationship is not obvious and it is possible that the figure is confounded by an independent effect of elevation or horizontal distance. Net of radial-bin means, the average sewer share inside the CBD basin and within 2km of the basin divide is 0.702, outside and within 2km is 0.655, and so the cross-divide difference is -0.046. Thus, our average treatment size is about 4.6 percentage points of sewer access.

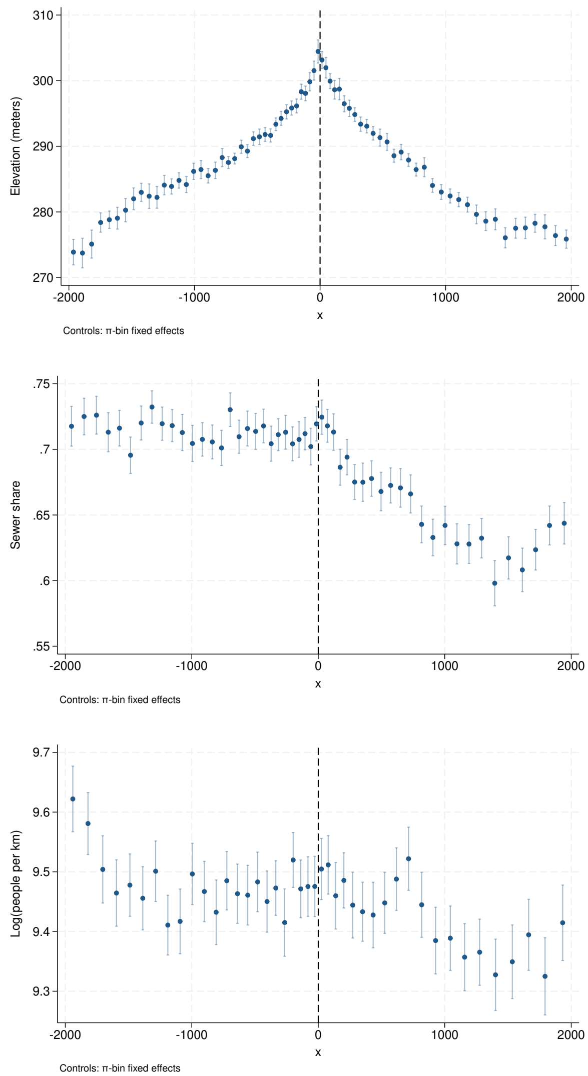
The bottom panel of figure 6 is like the middle panel, but reports bin means of log tract population density. This figure is less clear than the corresponding figure for sewer share, but population density appears to decline outside the central basin.

Our econometric specification exploits the variation illustrated in figure 6 by including an indicator for whether a tract is outside and the interaction of this indicator with horizontal distance to the basin divide. The validity of these instruments probably depends on controlling for elevation and horizontal distance, and we experiment with different specifications using these controls.

Figure 7 illustrates the empirical variation relevant to our second identification strategy. Like figure 6, both panels in figure 7 are binscatter plots. Like the middle and bottom panels of figure 6, the y-axis of the top and bottom panel of figure 7 reports bin means of sewer share and log tract population density net of radial-bin means. However, the x-axis of these figures is different from figure 6. In figure 7 the x-axis describes meters of climbing required to reach the central basin divide. Positive x values indicate meters of climbing to reach the basin divide for a tract on the outside of the basin, and negative x values indicate meters of climbing to reach the basin divide for a tract on the inside of the basin. For example, the mean sewer share is about 0.68 for tracts that are 100m below and outside the basin divide, while the mean sewer share is about 0.75 for tracts that are inside the basin divide and only a few meters below it. Looking carefully shows a clear break in sewer share at $x = 0$.

The bottom panel of figure 7 is like the top panel, but the y-axis reports the bin mean of log tract population density. This figure also shows a trend break and a step at $x = 0$.

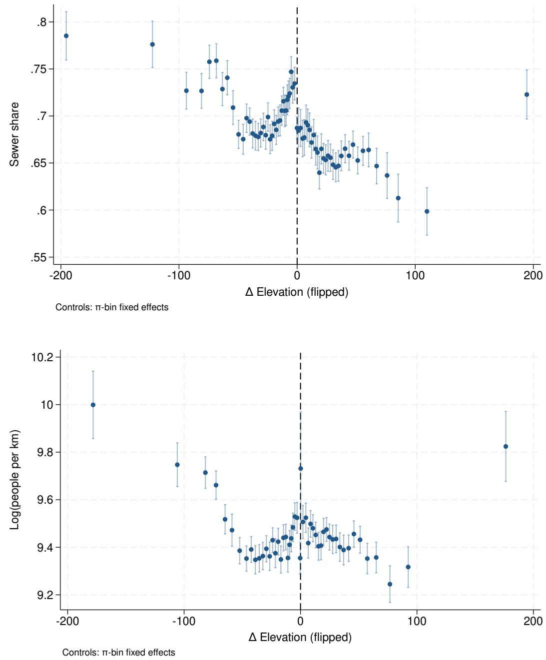
Figure 6: Identification, $z = \mathbb{1}(x \text{ is outside})$



Note: *All panels are binscatter plots. (Top) Mean tract elevation net of radial-bin fixed effects as a function of horizontal distance. (Middle) Mean tract sewer share net of radial-bin fixed effects as a function of horizontal distance. (Bottom) Mean tract population density net of radial-bin fixed effects as a function of horizontal distance. Top and middle are the empirical analogs to figure 1. All three panels are based on the main estimation sample described in column 2 of table 1.*

Our econometric specification exploits the variation illustrated in figure 7 by including an indicator for whether a tract is outside and the interaction of this indicator with vertical distance to the basin divide, while controlling for horizontal distance and elevation.

Figure 7: Identification, $z = \mathbb{1}(x \text{ is outside})\Delta h$



Note: *Both panels are binscatter plots. Top panel reports mean tract sewer share as a function of vertical distance to basin divide, net of radial-bin mean. Bottom panel is mean log tract population density as a function of vertical distance to basin divide, net of radial-bin mean. Both panels are based on the main estimation sample described in column 2 of table 1.*

7 Reduced form results

We begin by estimating the effect of the tract share of sewer access on log tract population density. Let j index census tracts and k index radial-bins. s_{jk} is the share, from zero to one, of households reporting sewer access in tract j and radial-bin k. y_{jk} is the outcome of interest, to begin the logarithm of population density. x_{jk} is meters from the tract centroid to the basin divide, with displacements inside the basin negative and displacements outside positive. $\Delta h_{jk} \geq 0$ is the vertical rise required to reach the basin divide from the centroid of tract j and $\mathbb{1}(\pi\text{-bin})_{jk}$ is an indicator that is one for all tracts in radial-bin k and zero otherwise. Finally, $\mathbb{1}(\text{Outside})_{jk}$ is an indicator that is one for tracts with centroids outside the central basin, and zero otherwise.

Our research design requires two estimating equations. The first is a first-stage predicting sewer share using all three (or a subset) of our instruments, $\mathbb{1}(\text{Outside})_{jk}$, $\mathbb{1}(\text{Outside})_{jk}x_{jk}$, and $\mathbb{1}(\text{Outside})_{jk}\Delta h_{jk}$, along with controls for radial-bin and elevation,

$$
\begin{aligned}
s_{jk} =& \mathbb{1}(\pi\text{-bin})_{jk} + \mathbb{1}(\pi\text{-bin})_{jk} x_{jk} + A^s \Delta h_{jk} + \\
& \alpha_0 \mathbb{1}(\text{Outside})_{jk} + \alpha_1 \mathbb{1}(\text{Outside})_{jk} x_{jk} + \alpha_2 \mathbb{1}(\text{Outside})_{jk} \Delta h_{jk} + \eta^s_{jk}.
\end{aligned} \tag{1}
$$

The second is a structural equation predicting a tract outcome as a function of tract sewer share and controls,

$$
y_{jk} = \mathbb{1}(\pi\text{-bin})_{jk} + \mathbb{1}(\pi\text{-bin})_{jk} x_{jk} + A \Delta h_{jk} + \beta s_{jk} + \eta_{jk}. \tag{2}
$$

Depending on estimation technique, equation (2) is an OLS estimation or a TSLS/IV regression.

Figure 5 demonstrates that the expected elevation profile around basin divides is only present once we control for radial-bin fixed effects. Because of this, all of our regressions include an indicator variable for each radial-bin.

Our three instruments are an indicator for outside, $\mathbb{1}(\text{Outside})$, this indicator interacted with the distance to the boundary, $\mathbb{1}(\text{Outside})x$, and this indicator interacted with meters of climbing required to reach the basin divide, $\mathbb{1}(\text{Outside})\Delta h$. We are concerned that horizontal distance has a direct effect on sewer share. Given this, we control for horizontal distance in three different ways. First, by including horizontal distance as a control. Second, by including horizontal distance interacted with segment-bin indicators, and third, by including the interaction of the radial-bin indicator with the horizontal distance, that is, $\mathbb{1}(\pi\text{-bin})_{jk} x_{jk}$. Because we also include a radial-bin indicator, in this final specification all of our parameter estimates are conditional on a radial-bin specific slope and intercept. We are also concerned that elevation has an independent effect on sewer share and outcome, and so we also control for elevation.

Table 2 presents estimation results for log population density. The top panel gives results of OLS regressions of equation (2) with different controls. Panel 2 presents first-stage regressions, equation (1), using different combinations of instruments and controls. Panel 3 presents TSLS estimates of equation (2) using the instruments and controls common to other results in the same column. We postpone discussion of panel 4. The bottom panel of the table describes the controls used in each specification, gives the sample size, and an F-statistic (Cragg-Donald) for the instruments in the first-stage regression.

Table 2: Sewers and log tract population density

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.8562***	0.8562***	0.8562***	0.8108***	0.8108***	0.8108***	0.7398***	0.7398***	0.7398***
	(0.0259)	(0.0259)	(0.0259)	(0.0265)	(0.0265)	(0.0265)	(0.0266)	(0.0266)	(0.0266)
2. First-stage									
Outside	-0.00800**	-0.00554	-0.00509	-0.01404***	-0.00335	-0.00448	-0.00984***	-0.00228	-0.00461
	(0.003561)	(0.003946)	(0.003946)	(0.003648)	(0.004385)	(0.004386)	(0.003773)	(0.004708)	(0.004716)
x*Outside	-0.00005***		-0.00005***	-0.00006***		-0.00006***	-0.00007***		-0.00007***
	(0.000004)		(0.000004)	(0.000005)		(0.000005)	(0.000006)		(0.000006)
ΔElev*Outside		-0.00018**	-0.00013		-0.00045***	-0.00040***		-0.00025**	-0.00021*
		(0.000076)	(0.000077)		(0.000099)	(0.000099)		(0.000110)	(0.000110)
3. IV									
Sewer share	1.8284***	3.5182**	2.0739***	3.9819***	2.5200***	4.1937***	6.0387***	0.4847	5.9413***
	(0.3559)	(1.4295)	(0.3591)	(0.4632)	(0.8744)	(0.4490)	(0.6072)	(1.5441)	(0.5890)
4. SATE									
Sewered	2.3389***	4.8942***	2.6074***	3.7429***	2.9344***	4.0004***	5.7074***	4.0916**	5.7297***
	(0.3641)	(1.2770)	(0.3662)	(0.3681)	(0.8500)	(0.3562)	(0.3850)	(1.4261)	(0.3774)
N	50805	50805	50805	50805	50805	50805	50805	50805	50805
F	92.34	7.061	62.70	81.16	20.15	60.65	73.25	5.924	50.32
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

The columns are in three groups of three. The first three columns (1-3) control for tract elevation, horizontal distance, and radial-bin intercepts. The second group of three (4-6) allows the effect of horizontal distance to vary by segment-bin, and the third group of three (7-9) allows the effect of horizontal distance to vary by radial-bin. The third group of three has the most flexible controls for potential confounding trends in horizontal distance and is our preferred specification.

Within each group of three columns, we vary the instruments that we use. In column 1 of each set (1,4,7), our instruments are the outside indicator and the outside indicator interacted with horizontal distance. In these regressions, causal identification relies on changes in elevation around the central basin boundary that results from horizontal distance.

In column 2 of each set (2,5,8) our instruments are the outside indicator and the interaction of vertical distance to the basin divide and the outside indicator. In these regressions, identification relies on changes in elevation around the central basin boundary holding horizontal distance constant. The coefficient on these variables tells us the amount by which sewer share or log population density decrease with each additional meter of climbing to reach the basin divide for tracts outside the central basin. The third column of each set of three (3,6,9) includes all three instruments.

In the top panel of table 2 we see that a 1 percentage point increase in sewer share is associated with a little less than a 1% increase in population density. This effect is estimated precisely and is stable across specifications. This about matches the pattern that figure A1 shows in the raw data. The OLS specification is identical for each of the three columns within a group, e.g., columns (1-3), because it is not affected by changes to the instrument set.

Panel 2 presents first-stage results. Point estimates for instruments have the expected negative signs. Whether measured by horizontal or vertical distance, being on the outside of the central basin divide increases the cost of sewer access and decreases its prevalence. The coefficients on the two interaction variables are stable across specifications and are estimated precisely. The relevant F-statistic is above the threshold for conventional weak instrument tests, except in columns 2 and 8. We discount estimates based on these first-stage estimates, although we note that weak instrument tests only apply to models with homogeneous treatment effects.

Columns 2 and 8 both use the outside indicator and the outside indicator interacted with vertical distance to the divide as instruments. Figure 5 suggests an explanation for why the elevation variable does not explain much of the variation in sewer share; there is just not very much variation in elevation once we condition on horizontal distance to the basin divide. This intuition is also consistent with the results in columns 2 and 5 of table 2. When we allow the coefficient on horizontal distance to vary over larger sections of the basin divide (so there is more variation in elevation holding distance to the divide constant) first stage explanatory power increases.

Finally, panel 3 presents our TSLS results. Ignoring columns 2 and 8, estimates range between about 1.8 in column 1 and 6.0 in column 7. Estimates generally increase as we add controls. Column 9 has the most exhaustive set of controls and is our preferred estimate at 5.94. Comparing the TSLS elasticities of panel 3 with the OLS elasticities in panel 1, suggests that the causal effect of sewer access is between about two and six times as large as the correlation in the raw data.

Conceptually, our research design involves randomization of treatment at the level of the census tract. This means that treatment assignment occurs at the level of the unit of observation. Thus, the standard rationale for clustering standard errors does not apply. With this said, tract elevation is calculated in part on the basis of the radial-bin maximum elevation, and this creates the possibility for within radial-bin error correlation. To address this, table A2 repeats table 2 but reports standard errors clustered at the level of the radial-bin. We see that clustered errors are two to three times as large as robust errors, but still indicate that we measure treatment effects precisely.

This requires two comments. First, tables 2 and A2 reveal two patterns that occur across almost all of the outcomes and specifications that we present. First, that causal estimates of treatment effects are larger than OLS. This indicates that the equilibrium assignment process tends to provide sewer access to places that have smaller outcome values (population density in the current case). This seems consistent with the provision of sewers in newer and lower density development on the edges of cities. Second, we present robust standard errors in our main results to allow comparison with the SATE estimator presented in panel 4 of table 2 for which clustered errors would be difficult to calculate. We have calculated clustered standard errors for all of the TSLS results that we present, and except for a few cases that we note, they follow the pattern that we see when we compare tables 2 and A2. That is, clustered errors are larger than robust errors, but still absolutely small.

Appendix tables A3, A4, and A5 repeat the estimates of table 2 separately for, Brazil, Colombia, and South Africa. In all three cases, the first stage is qualitatively similar to what we see in the whole sample. The first stage is strong for Brazil and Colombia, but weak for South Africa, where our sample is much smaller. Focusing attention on our preferred specification in column 9, for Brazil and Colombia, treatment effects are 5.6 and 4.7. For South Africa, the sign of the effect flips. Places with greater sewer access are on average less dense than those without. The TSLS estimate of this effect is imprecise, but we see the same change of sign in the OLS estimate as well. Our samples for Tanzania and Jordan are too small to permit the estimation of our TSLS specification, however, the OLS estimate of the effect of sewers on population density and is negative for Tanzania and positive for Jordan. These results suggest heterogeneous treatment effects.

8 Instrument validity

We can imagine a number of challenges to our identification strategy. First, by construction, each central basin encircles the center of the city it contains. Thus, we expect that displacement outside is away from the city center and leads land to become less valuable and less intensively developed for a reason unrelated to a tract's location relative to the basin divide.

We have three responses to this problem. First, we note that the relationship between distance to the CBD is weaker than might be expected. In a regression of horizontal distance on distance to the CBD, the R^2 is approximately zero. Repeating this regression with the addition of radial-bin fixed effects, the within-R^2 is still only 0.37. Second, we control for horizontal distance to the divide in a flexible way. In our preferred estimates in columns 7-9 of table 2 we include both radial-bin specific intercept, and a radial-bin specific slope in horizontal distance. To the extent that radial distance from the CBD and horizontal distance from the basin divide are correlated, we are controlling flexibly for radial distance. Third, we control explicitly for radial distance in table A6. These results are similar to our main results. In all, the problem of confounding distance to the CBD with horizontal distance appears not to be important.

Because it is impractical to operate a sanitary sewer without a piped water supply, piped water and sewer are sometimes installed simultaneously. This raises the possibility that our estimates reflect their joint effect. There are a number of reasons to think this is not the case.

First, residential piped water is not a gravity fed system and is typically delivered at a pressures between 40-80 psi. This is more than enough pressure to deliver water over an average basin divide.[7] Thus, piped water delivery is not affected by the sorts of elevation changes typically found in a neighborhood of a basin divide except to the extent that it is jointly provided with sewer access. Second, because piped water is easier to provide than sanitary sewers, it is more common. This means that we can restrict our estimation of sewer access treatment effects to tracts where piped water access is universal. In appendix table A7 we duplicate table 2 restricting attention to the sample of tracts where

[7] A one inch square water column, 100 feet high, weighs a little over 40 pounds.

piped water access is universal. This sample restriction reduces the coefficient on sewer share from 5.9 in our main specification to 5.0 in the constrained sample, and the precision of our estimates does not allow us to reject the hypothesis that the two are the same. Finally, in appendix table A8 we replicate the results of table 2 adding a control for the tract share of piped water access. This increases the point estimate on sewer share to about 6.8. Summing up, while piped water access is sometimes installed jointly with sanitary sewers, this does not appear to confound our estimation of the effects of sewers on population density.

There is also evidence that municipal garbage collection is important for the operation of sanitary sewers. Cole et al. (2025) document that, absent municipal garbage collection, garbage ends up in sanitary sewers and blocks them. Thus, garbage collection is likely an input into the provision of sanitary sewer service in much the same way as piped water. This raises the possibility that our estimates reflect the joint effect of piped water and garbage collection. There are a number of reasons to think this is not the case.

First, there is no reason to think that the variation in elevation around a basin divide has any affect on garbage collection, except to the extent that garbage collection is bundled with sewer access. Second, as with piped water, garbage collection is more common than sanitary sewers, so that we can restrict our estimation of sewer access treatment effects to tracts where garbage collection is universal. Appendix table A9 replicates our main results after restricting attention to tracts where garbage collection is universal. In our preferred estimate of column 9, the point estimate for sewer share falls to about 4.0. Finally, we control for garbage collection in the regressions presented in appendix table A10. Following the pattern we saw for piped water, this increases the point estimate for the sewer share effect to about 6.7. Together these results suggest that even though garbage collection is sometimes provided jointly with sanitary sewers, we are probably not confounding their effects.

It is possible that the basin boundary is an important geological feature and that crossing it affects sewer share and outcomes because it impedes the movement of people and goods along with wastewater. The top panel of figure 6 (and the bottom panel of figure 5) suggests that this is rarely the case. Traveling 2km inside from an average basin divide involves a drop of only about 30m, an average grade of about 1:70. This is suggestive, but could conceal dramatic local variation. The whiskers in both figures describe 95% CIs of each distance bin mean. That these confidence intervals are so tight suggests that local elevation profiles seldom differ much from the average. Thus, the raw data suggest that basin divides dramatic enough to impede the movement of goods and people are rare.

To investigate the possibility that our results are driven by rare radial-bins where the basin divide is a dramatic geological feature, appendix table A11 replicates table 2 excluding radial-bins that contain a tract whose centroid is more than 100m below the basin divide. Relative to table 2, point estimates of TSLS treatment effect fall slightly in all specifications. However, in no specification does the change of sample result in a large enough change to coefficients that we fail to reject the hypothesis of no difference at standard confidence levels.

It is also possible that jurisdictional boundaries follow basin divides. If so, then crossing basin divides could affect sewer provision and population density because it involves crossing from one administrative unit to another. In our census data we can impute each tract's municipality from its census identification code, and then check if municipal boundaries tend to follow basin divides.[8] To test the extent that municipal jurisdictions follow basin boundaries, we identify radial-bins where any of the tracts within 250 meters of the basin divide are in different municipalities. In our estimation sample only 3.5% of the radial-bins (containing 5.6% of tracts) contain municipal boundaries close to the basin divide. Basin divides rarely coincide with municipal boundaries.

To investigate the possibility that our results are driven by the rare radial-bins where municipal boundaries and basin divides are close to each other, in appendix Table A12 we replicate table 2 excluding the 3.5% of radial-bins where any of the tracts within 250 meters of the basin divide are in different municipalities. Point estimates of TSLS treatment effects fall slightly in all specifications. However, in no specification does the change of sample result in a large enough change to coefficients that we fail to reject the hypothesis of no difference at 5% confidence.

[8]For example, in Colombia the first 5 digits of the identification code for a *manzana* or census block is that manzana's municipal identification code. Similarly, for Brazil, Jordan, South Africa, and Tanzania, we retrieve a tract's administrative unit from the census information provided for each country.

Two placebo tests also support our identification strategy. Figure A2 is a binscatter plot showing mean distance to the CBD net of radial-bin means. As expected, this plot is continuous and smooth. This is intuitive, but not guaranteed. If we had seen a step or kink in this distance gradient when we crossed the basin divide, it would have indicated a problem. Second, appendix table A13 repeats table 2, changing the outcome to share of tract households with electric service. There is no reason that the cost of electricity service should respond to crossing a basin divide, and so no reason to expect a causal effect of sewers on electricity access. Point estimates indicate that increasing sewer share from zero to one increases electricity share by between 2% and 10% and is generally decreasing as we add controls. In our preferred estimate of column 9, the relevant point estimate is 3% and is not distinguishable from zero at conventional significance levels.

Finally, in theory, it is also possible that sewers discharge close to the households they serve. If so, the beneficial effects of sewers may be confounded by the harmful effects of proximity to an untreated sewer outflow. In practice, this seems unlikely to be a problem in our sample. By construction, basin divides tend to be remote from the river or stream that drains the basin. Because sewer outflows, treated and not, are usually into waterways, this means that our sample households should usually be far from sewer discharges.

9 Estimating treatment effects from aggregate data

The econometric model described by equations (1) and (2) is an instrumental variables estimation with parametric controls and a continuous treatment. Angrist et al. (2000) and Kolesár and Plagborg-Møller (2024) study causal interpretation of linear TSLS with a continuous treatment and heterogeneous treatment effects. They show that the TSLS estimand can be interpreted as a weighted average of the marginal effects, although this average is difficult to interpret and not of obvious economic interest. Furthermore, the weights can be negative for some specifications and assumptions about the first-stage equation.[9] As a practical matter, this means that we can only treat our estimates of equations (1) and (2) as an average of treatment effects if we are willing to assume that treatment effects are homogeneous.

The results in appendix tables A3, A4 and A5 argue against an assumption of homogeneous treatment effects. Similarly, variation in the TSLS estimates in table 2 is consistent with a heterogeneous treatment effects model where different instruments lead to different weighted averages of treatment effects. Thus, our results so far suggest heterogeneous treatment effects, and therefore indicate the need for an alternative to TSLS.

To develop such an alternative, we note that sewer access is binary at the parcel level. A parcel has access or it does not. Thus, if we can estimate our model at the parcel level, we resolve the problem of continuous treatment. With the problem of a continuous treatment out of the way, we can use the MTE model of Carneiro et al. (2011) which permits the estimation of heterogeneous treatment effects in instrumental variables models with parametric controls and a binary treatment. Putting this all together, if we can estimate the MTE model at the parcel level from the available tract level data, the resulting estimates can be used to calculate any of the conventional (parcel level) average treatment effects using methods developed in Carneiro et al.

The obvious obstacle is that our data describe census tract means, not the underlying parcel level observations. We resolve this problem with a small variance approximation of the parcel level MTE model. Loosely, we write the parcel level MTE model, and take expectations over tracts to arrive at a tract level model. This leads to tract level estimating equations that identify the causal effects defined in the parcel-level MTE model on the basis of tract level aggregates. Because we take tract level expectations of non-linear parcel level

[9] Related to this, Słoczyński (2021) and Blandhol et al. (2022) both investigate the properties of the TSLS estimator with linear additive covariates and binary treatment. Both conclude that this model leads to a weighted average of treatment effects that allows a causal interpretation only under restrictive conditions. Indeed, Blandhol et al. (2022) argues that these conditions are so strict as to be impossible to satisfy in practice. Moreover, our problem involves a continuous (not binary) treatment. Whereas the binary case requires that we consider only two potential outcomes, treated and not, with a continuous treatment, we must consider a continuum of counterfactual outcomes for each unit. Chesher (2003) and Imbens and Newey (2009) consider the problem of estimating causal effects in a model with a continuous treatment, although their approaches are challenging when multi-dimensional covariates are present.

equations this leads to tract level estimating equations that depend on the within-tract variances of parcel level characteristics. Our data allows the calculation of these variances, and so allows us to estimate tract level averages of parcel level treatment effects.

We begin by stating the parcel level MTE model. Let i index parcels and j index census tracts. Y_{ij} is the outcome variable of interest for parcel i in tract j, D_{ij} is our treatment variable and takes the value one if a parcel has sewer access, and zero if not. X_{ij} is a vector of covariates, and Z_{ij} a vector of instruments. Let $W_{ij} = (X_{ij}, Z_{ij})$. Controls and instruments are the same as in the earlier TSLS estimates of equations (1) and (2).

Our parcel level MTE model consists of two linear equations describing the relationship between controls and potential outcomes,

$$Y_{ij}(1) = X'_{ij}\beta_1 + U_{ij}(1),$$
$$Y_{ij}(0) = X'_{ij}\beta_0 + U_{ij}(0),$$

along with a parcel level selection equation,

$$D_{ij} = 1\{p(W_{ij}) \geq V_{ij}\}.$$

As is standard in the MTE literature, the selection equation assumes additive separability of the unobserved heterogeneity, V_{ij}, and the latent utility term involving W_{ij}. Moreover, and without loss of generality, we normalize the unobserved parcel level heterogeneity in the selection equation, V_{ij}, to be uniform on the unit interval. As a result, $p(W_{ij})$ coincides with the parcel level propensity to select into treatment $\Pr(D_{ij} = 1|W_{ij})$.

We also impose the practical exogeneity condition of Carneiro et al. (2011), $(U_{ij}(1), U_{ij}(0), V_{ij}) \perp W_{ij}$. This implies that unobservable heterogeneity in the parcel level selection and potential outcome equations is statistically independent of parcel level observables. In this case, the local IV regression is given by

$$Y_{ij} = X'_{ij}\beta_0 + p(W_{ij})X'_{ij}(\beta_1 - \beta_0) + \phi(p(W_{ij})) + U_{ij}, \qquad (3)$$

where $\phi(\cdot)$ is a control function describing the dependence between $(U_{ij}(1), U_{ij}(0))$ and V_{ij}. The parcel level marginal treatment effect (MTE) conditional on $X_{ij} = x$ and $V_{ij} = v$ is the derivative of equation (3) with respect to the propensity score. That is,

$$\text{MTE}(x,v) \equiv E[Y_{ij}(1) - Y_{ij}(0)|X_{ij} = x, V_{ij} = v] = x'(\beta_1 - \beta_0) + \phi'(v). \tag{4}$$

A consequence of V_{ij} uniform and the practical exogeneity condition is that the Conditional Average Treatment Effect given X_{ij} (CATE) is linear in X_{ij},

$$\text{CATE}(X_{ij}) = E[\text{MTE}(X_{ij},V_{ij})|X_{ij}] = X'_{ij}(\beta_1 - \beta_0) + \int_0^1 \phi'(v)dv. \tag{5}$$

In words, $\text{CATE}(X_{ij})$ gives the treatment effect for an average parcel with observable characteristics X_{ij}.

Our problem is to estimate this model using tract level data. We begin by introducing notation to describe tract level means and variances of parcel level data. Let $\sigma^2 R_j$ denote the variance-covariance matrix of W_{ij} within each tract j, with σ^2 a scaling term that we introduce to facilitate our small-variance approximation. Our data reports tract averages, $(\overline{Y}_j, \overline{D}_j, \overline{W}_j)$, and $\sigma^2 R_j$. We discuss the calculation of $\sigma^2 R_j$ below.

To estimate a parcel level model from tract means, $(\overline{Y}_j, \overline{D}_j, \overline{W}_j)$, and the within tract variances of W_{ij}, $\sigma^2 R_j$, we make three additional assumptions. First, that the parcel level propensity score $p(W_{ij}) = \Pr(D_{ij} = 1|W_{ij})$ is three times continuously differentiable. Second, that the third moments of W_{ij} exist. Third, that $\phi(p)$ is quadratic,

$$\phi(p) = \alpha_0 + \alpha_1 p + \frac{1}{2}\alpha_2 p^2. \tag{6}$$

Given these assumptions, Appendix C derives a tract-level regression equation that estimates the parcel level MTE model by using a small variance decomposition.

If we further assume a linear probability model for the propensity score,

$$p(W_{ij}) = W'_{ij}\gamma. \tag{7}$$

then we can estimate the parameters $(\beta_1 - \beta_0)$, α_0, α_1, and α_2 from the parcel level first-stage and structural equations (3) and (7), with the following two tract level estimating equations,

$$p(\overline{W}_j) = \overline{W}_j'\gamma, \tag{8}$$

$$\begin{aligned}\overline{Y}_j = \overline{X}_j'\beta_0 + p(\overline{W}_j)\overline{X}_j(\beta_1 - \beta_0) + \sigma^2(\beta_1 - \beta_0)k_{1j} \\ + \alpha_0 + \alpha_1 \cdot p(\overline{W}_j) \\ + \frac{1}{2}\alpha_2 \cdot [p(\overline{W}_j)^2 + \sigma^2 k_{3j}] + O_j(\sigma^3) + \eta_j,\end{aligned} \tag{9}$$

where k_{1j} and k_{3j} are observable scalars calculated from $\overline{W}_j$, R_j and $\widehat{\gamma}$, and $O_j(\sigma^3)$ is a tract level approximation error that vanishes as $\sigma^3 \to 0$. These are the original MTE structural equations estimated on tract averages, but with the addition of terms involving the variances of tract level variables in the structural equation.

Equation (9) is linear in parameters, even though some of the regressors involve non-linear calculations. Hence, assuming that the approximation error term $O_j(\sigma^3)$ is negligible, we can estimate equations (8) and (9) with OLS to identify $(\beta_1 - \beta_0)$, α_0, α_1, and α_2. This lets us estimate $\text{MTE}(x,p)$ and other causal estimands by plugging these parameters into equations (4) and (5).

We obtain the tract level sample average treatment effects by averaging $\text{CATE}(X_{ij})$ over all parcels in each tract j. Under the practical exogeneity assumption, this leads to an expression for the sample average treatment effect for a tract with mean parcel observables $\overline{X}_j$,

$$\begin{aligned}\text{SATE}_j &= \overline{X}_j'(\beta_1 - \beta_0) + \int_0^1 \phi'(v)dv \\ &= \overline{X}_j'(\beta_1 - \beta_0) + \alpha_1 + \frac{1}{2}\alpha_2.\end{aligned} \tag{10}$$

We report the average of this quantity over all sample tracts in panel 4 of table 2 and several of our other tables. We discuss the calculation of standard errors in Appendix C.

Recall that our identification strategy relies heavily on radial-bin dummy variables and their interaction with the horizontal distance to the basin divide. With so many regressors, our estimators are difficult to compute, particularly if we rely on a non-linear, e.g., Logit, functional form for the propensity score $p(\cdot)$. Moreover, the incidental parameter problem arises in the estimation of fixed effects in nonlinear regressions (see e.g., Lancaster (2000)). These two problems motivate our reliance on the linear probability model described by equation (8).

That we rely on a large set of controls creates another problem. The statement of the MTE model allows for treatment effects to vary with unobserved resistance to treatment and with observed characteristics, the term $p(W_{ij})X'_{ij}(\beta_1 - \beta_0)$ in equation (3). Given the large number of fixed effects in our regression equations, this specification is an extremely flexible description of treatment heterogeneity, and in practice, estimates of the $(\beta_1 - \beta_0)$ are difficult to compute and unstable, and so we impose the restriction $(\beta_1 = \beta_0)$. This restricts attention to the case where treatment effects are heterogeneous on unobservables only.

To estimate this model, we require data describing the within tract variances of parcel level control variables and instruments. The list of instruments and controls that we use in our estimates of equation (2) involves four variables and interactions of these variables; elevation, horizontal distance, an outside indicator, and radial-bin indicators. By construction, radial-bin indicators are constant within a tract, and so their within tract variance is trivially zero. This leaves elevation, horizontal distance, and the outside indicator.

Our census data does not report any data at the parcel level. However, our elevation data is gridded data with a spatial resolution of about 30 meters, about the size of a parcel. We can evaluate the outside indicator and horizontal distance for each grid cell in the elevation data. We also have tract boundary files. Putting these data together, we can calculate the within tract variances and covariances of all of our control variables and instruments for the universe of tracts in our study area. To implement the estimator described by equations (8) and (9), we use within tract calculations of variances based on 30m grid cells to proxy for parcel values, and estimate both equations with OLS.

With these estimates in hand, we can evaluate the SATE_j described in equation (10) for each tract in our sample. Averaging over the sample of tracts we estimate for the average treatment effect.

The resulting estimates are reported in panel 4 of table 2, along with standard errors calculated as we describe in Appendix C. Ignoring columns 2 and 8 where the instruments are weak, the range of SATE estimates is from about 2.3 to 5.7, narrower than the about 1.8 to 6.0 range of estimates for TSLS. The SATE estimates and the TSLS estimates are close compared to the precision of the estimates, particularly for our preferred specification in column 9, and the SATE estimates are marginally more precise than the TSLS estimates. Unlike the TSLS estimations, the SATE estimate in column 8 is not an outlier. This is what we would expect if the different TSLS estimates were weighted averages of different marginal treatment effects for different sets of compliers, but the SATE correctly evaluates the same average treatment effect.

While it is tempting to interpret TSLS coefficients as LATES, this interpretation rests on strong assumptions, and the resulting regression weighted LATE is not obviously of economic interest. SATE, on the other hand, has a straightforward interpretation. It describes the amount by which population density changes in an average parcel when that parcel receives sewer access.

10 Other outcomes

We next investigate whether sewers affect the demographic characteristics of tract residents. This is of intrinsic interest and helps us to understand the incidence of the benefits of sewer access. Do sewers help incumbent residents, or do they precipitate the arrival of more affluent migrants?[10]

We investigate this question in two ways. First, the Brazilian and South African censuses report tract income directly. For these two countries, we use the same procedure that we used to investigate the effect of sewers on density. Appendix table A15 reports results. This table is identical to table 2, except that the dependent variable is the log of tract mean income.

Recalling that the regressor of interest is a share and the dependent variable is a logarithm, the OLS results indicate that a 1 percentage point increase in sewer share is associated with about 0.4% increase in income. Estimates in panels three and four are somewhat larger. In our preferred estimate in column 9, a 1 percentage point increase in sewer share causes about 0.8% increase in tract mean income for the TSLS estimations and about 0.7% for the SATE estimate. The precision of the TSLS and LATE estimates is marginal, and if we cluster errors at the radial-bin level, as in appendix table A2, the standard errors on the OLS and TSLS estimates increase to 0.03 and 0.94. In this case, we cannot distinguish the TSLS coefficient from zero at conventional significance levels.

[10]Looking at the effects of sewers on mortality is also of obvious interest. We experimented with measures of infant and childhood mortality constructed from census count data, but found that they were too noisy to be informative. This inquiry appears to require vital statistics data, and is a subject for future research.

From table 1, we see that the mean and standard deviation of income are both about 950 dollars. Abusing the marginal nature of our result, providing universal sewer access to a previously unsewered tract increases income by a factor of $e^{0.82} \approx 2.2$. Thus, providing universal sewer access to a previously unsewered tract increases tract mean income by about 1.2 standard deviations. This about doubles tract mean income from 933 to 1930USD2022. Two caveats are required here. First, we estimate this effect imprecisely, and so it could well be smaller. Second, most tracts in our sample are partly sewered, so interventions this large will be rare in practice.

We next investigate whether sewers affect the literacy rate of tract residents. Appendix table A14 reports results. OLS estimates suggest the literacy rate is about 0.02 percentage points higher in a tract when the sewer share is 1 percentage point higher. IV results, in panels three and four, suggest a larger effect. In our preferred specification of column 9, a 1 percentage point increase in sewer share causes about a 0.04 percentage point increase in the literacy rate, although this effect is not distinguishable from zero. If we consider standard errors clustered at the radial-bin level, then the standard error for the OLS estimate of column 9 increases to about 0.004 and for the TSLS estimate, to about 0.05. The first is well below conventional significance thresholds, and the second well above. Even if we take the 0.04 percentage point IV estimate of column 9 seriously, providing universal access to a completely unsewered tract increases the literacy rate by about 4 percentage points. From table 1 the tract mean literacy rate is about 92% with a standard deviation of about 40%. This suggests that sewer access plays at most a small role in the spatial distribution of literacy.

In all, our results suggest that improved sewer access precipitates the arrival of migrants who are modestly better off than incumbents, but these migrants will be at most slightly more literate.

We now consider the effect of sewers on other characteristics of the built environment. As we discussed earlier, piped water and garbage collection are inputs to the provision of sewer service. Thus, we expect sewer provision to sometimes require improvements to these services as well. Appendix tables A18 and A17 establish that this occurs. These two tables repeat the analysis of tract population density in table 2, but using the share of tract households with piped water and garbage collection as outcomes. Focusing on our preferred specification in column 9, each one percentage point increase in tract sewer access causes about a 0.5 percentage point increase in tract household piped water access and 0.4 percentage point increase household garbage collection. This suggests that whenever sewers are installed and the required inputs, piped water and garbage collection, are not in place, they are also made available.

These results require two comments. First, recalling appendix tables A8 and A9, the effect of sewers on tract population density appears to result from the provision of sewers, not piped water or garbage collection. Second, sewer access does not appear to lead to a more general upgrading of public services. Among the public services we can check, piped water, garbage collection and electricity, sewers cause increases only to the two inputs to sewer service, and not to electricity service (which does not play a role in the provision of sanitary sewers).

Finally, we consider the effects of sewer provision on the share of households that report living in an apartment. Appendix table A16 replicates the analysis of table 2 using the share of households living in an apartment as the outcome. Ignoring columns 2 and 8 where the first stage is weak, both TSLS and SATE estimates range between about 0.37 and 0.49 and are estimated precisely. OLS estimates are also consistently positive, about 0.09, and are also estimated precisely. If we calculate errors clustered at the radial-bin level, errors on the TSLS estimates increase by about a factor of three. This puts the estimates in columns 7 and 9 just above the 5% significance threshold, although columns 1, 3, 4 and 6 remain below it. In all this provides reasonably strong evidence that part of the way that sewered tracts accommodate more people is by building bigger buildings.

11 Discussion

Sewers and urban density

Using our preferred estimate from column 9 of table 2, adding 1% of sewer connections to a tract causes an increase in tract population density of about 6%. It is not immediately clear whether this is an economically important effect. We would like to develop some intuition around this issue.

For each city in our sample, consider a counterfactual case where we add 1%

to the count of sewer connections in the city.[11] We add these connections, tract by tract, by first sewering all unsewered households in the most densely populated tract where sewer access is not universal. If completing sewer coverage in this tract does not exhaust the 1% increase in total connections, we move on to the next most densely populated tract containing unsewered households, and so on, until we allocate all of the 1% of new connections.[12]

For each city, this process results in a counterfactual city where a subset of tracts has better sewer access than in the observed case. We can then use our estimates of treatment effects to calculate the implied increase in population in these tracts. Assume that the 1 percentage point increase in sewer connections increases city population by inducing rural residents to migrate to the city, and a 6% treatment effect. In this case, mechanically, our counterfactual cities house 6% more people than their observed counterparts.

This effect seems large in the following sense. Baum-Snow (2007) finds that each radial interstate highway decreased the density of US central cities by 9%. Our estimates suggest the opposite effect can be accomplished by adding about 1.5% to a central city's stock of sewer connections. That is, a 1% increment to a central city's sewer share is about two thirds as important for urban form as is a single limited access radial highway.

The increase in person weighted density is also of interest, but must be evaluated tract by tract. We perform this calculation for all cities in our estimation sample. The left panel of figure 8 presents a histogram summarizing our results. The median city in our sample experiences an about 16% increase person weighted density, and the upper tail experiences much larger increases.

These effects seem large. The relationship between density and labor productivity is well established, and a central estimate is that doubling the density of a city increases wages by about 5%. Combining this estimate with the median 16% increase in person weighted density from figure 8, suggests that

[11] For this purpose, a city consists of a set of census tracts with the following properties. First, they are all contained in the central basin, part of our main estimation sample, or within 4km of the city center. Second, if this population exceeds the UN population for the city, we drop tracts until we are under this threshold, starting from the most remote and working inward.

[12] For reference, 75% of households have sewer access in an average tract in the sample described by column 1 of table 1. On the other hand, in the discussion surrounding figure 6, we saw that our average treatment size is about 4.6%. Thus our counterfactual exercise applies our estimates to treatments about 5 times as large as those on which our estimates are based.

adding 1% to the stock of sewer connections will increase wages of incumbent residents by about 0.80%. This is a flow. Taking its discounted present value using a 5% interest rate gives about 16% of the city's total annual wage bill. From table 1, income in an average tract (in South Africa or Brazil) is 968USD2022 per month or 11,616USD2022 per year. Mutiplying, we have that this sewer expansion increases the discount present value of urban income by about 1860USD2022.

Mara (1996) estimates the construction cost of simplified sewerage[13] at 22-34USD2005 per person. Multiplying by four people per household and adjusting to 2022 dollars gives a range from 133-207 per household. (Selendy, 2011, Ch2, table 2.4) estimates construction costs per household of 52-112$ for simplified sewerage or 120-160$ for conventional sewerage around 2000. Updating to 2022 dollars, we have 86-185 and 198-264 , respectively.[14] (Hutton et al., 2007, Table 4) gives a range of 22-25USD2007 per person for water and conventional sewer access. Again multiplying by 4 and updating to 2022 dollars, this is 134-152 . All of these cost estimates are an order of magnitude smaller than the wage increase calculated above. Adding the benefits that accrue to city residents from reduced morbidity and mortality, and the benefits to urban migrants, would make the cost-benefit comparison even more lopsided.

CBD access and sewer access

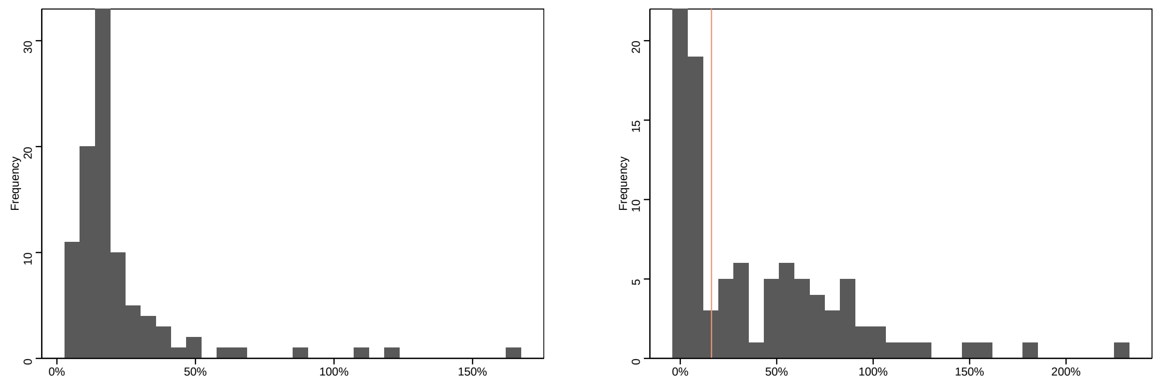
It is now common to evaluate public transit systems at least partly on the basis of the extent to which they improve access to the central city and thereby improve the functioning of the labor market, e.g., Tsivanidis (2019), Zárate (2022), Heblich et al. (2020). By facilitating higher residential densities, improving the sewer network within walking distance of the CBD also improves access to the CBD.

To assess the importance of this effect, we ask how many people would gain access to the CBD if we completed the sewer network in all tracts with centroids within 4KM of the CBD, and then allowed density to adjust. Using our same 6%

[13]"Simplified sewerage" is constructed using smaller drains and requires somewhat more maintenance than the sewers typically constructed in developed countries. It is also much less expensive (Mara, 1996).

[14]Price adjustments calculated using https://download.bls.gov/pub/time.series/cu/cu.data.0.Current, accessed December 2025.

Figure 8: Histograms of counterfactual changes in density and percentage with access to CBD



Note: *Left panel: Histogram of mean increase in person weighted density by city that results from adding 1% more sewer connections to the most densely populated tracts where access is not universal. Right panel: Histogram of showing the change in the percentage of city population that would gain walking access to the* CBD *if all tracts within 4KM of the* CBD *had sewer access increased to 100% and the consequent population increase was from more remote tracts. Vertical red line at 18% is the share of Bogota's population that gained access to the* CBD *because of the Transmilenio* BRT *system.*

treatment effect, we find that completely sewering the area within 4KM of the CBD increases the population within walking distance of the center by 41% of the city total for the average city in our sample.

This 41% mean value conceals cross-city heterogeneity. The right panel of figure 8 is a histogram of describing the magnitude of the effect. From Tsivanidis (2019), the Transmilenio BRT allows about 18% of the population of Bogota to access the CBD. Figure 8 suggests that building out the central city sewer network has an effect equal to or larger than that of a successful BRT system in 54% of our sample cities.

12 Conclusion

Few infrastructure investments seem as likely to have an important effect on the layout and growth of cities as the provision of sewers. However, sewers may increase density by resolving problems associated with on-site management of sewage, but by making a place nicer, they may draw in wealthier residents who demand more space. Similarly, it is ambiguous a priori whether sewer construction precipitates a general improvement in public services and amelioration of slum conditions, or if it has more limited follow on effects. We begin to answer these questions.

We estimate that a 1 percentage point increase in the share of tract households with sewer access causes about a 6% change in tract population. This effect is estimated precisely and is similar for the both TSLS and SATE estimators. There is suggestive evidence for heterogeneous effects across countries. While the effects of sewers on density in Brazil, Columbia and Jordan are close to the population average, sewer access probably decreases density in South Africa, and possibly in Tanzania.

We perform two simple counterfactual experiments to assess whether our estimated 6% effect size is economically important. The first of these suggests that adding 1% to a city's sewer connections in its densest neighborhoods has an effect on population density about 2/3 as large and of opposite sign as as a single radial interstate highway ray. The second experiment suggests that completely building out the sewer network within 4km of the CBD will increase the share of city population living in this disk by 40% on average, and by much more in some cities. This means that building out central city sewer access is often as important for improving access to central city labor markets as is a successful BRT system. A back of the envelope calculation suggests that if we consider only the increase in labor productivity following from increased density, the benefits of sewer construction in central areas outweighs their costs by about a factor of ten.

Piped water access and garbage collection are inputs into sanitary sewer service. Unsurprisingly, we find that in an average tract, the share of households with piped water access and garbage collection increases by about half as much as does the increase in the share with sewer access. This is consistent with the expansion of piped water access and garbage collection whenever sewers are extended into places where these other services are not provided. Also unsurprisingly, the share of households living in apartment buildings appears to increase with the construction of sewers. With this said, improvements to sewers do not lead to expansions of public services that are not complementary to sewer access. Electricity provision does not respond to change in sewer access.

The effects of sewer access on the demographic characteristics of tract residents seems small. We cannot distinguish the effect of sewer access on tract literacy rates from zero. The effect of sewer access on the logarithm of tract mean income is small statistically. With robust errors, we only just distinguish the effect of sewers on income from zero at conventional significance levels, and have even less precision if we cluster errors at the radial-bin level. Taking our point estimates seriously, providing universal sewer access to a previously unsewered tract about doubles tract income on average, from about 950$ to about 1900$ per month.

Because the treatment we consider, share of tract households with sewer access, is continuous, the interpretation conventional TSLS estimates is difficult except under strong assumptions. To arrive at an estimand that can be interpreted as an average treatment effect in a heterogeneous treatment effects framework, we also estimate a Sample Average Treatment Effect. In practice, both approaches lead to similar estimates of effect size. In addition to our contribution to understanding the effects of sewers on urbanization, we hope that this technique will prove useful to other researchers faced with similar estimation problems.

References

Aldous, D. (1999). *International turf management handbook*. CRC Press.

Allcott, H., Collard-Wexler, A., and O'Connell, S. D. (2016). How do electricity shortages affect industry? evidence from india. *American Economic Review*, 106(3):587–624.

Alsan, M. and Goldin, C. (2019). Watersheds in child mortality: The role of effective water and sewerage infrastructure, 1880–1920. *Journal of Political Economy*, 127(2):586–638.

Anderson, D. M., Charles, K. K., and Rees, D. I. (2018). Public health efforts and the decline in urban mortality. Technical report, National Bureau of Economic Research.

Angrist, J., Graddy, K., and Imbens, G. (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Review of Economic Studies*, 67(3):499–527.

Ashraf, N., Glaeser, E., Holland, A., and Steinberg, B. M. (2017). Water, health and wealth. Technical report, National Bureau of Economic Research.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2):775–805.

Bhalotra, S. R., Diaz-Cayeros, A., Miller, G., Miranda, A., and Venkataramani, A. S. (2021). Urban water disinfection and mortality decline in lower-income countries. *American Economic Journal: Economic Policy*, 13(4):490–520.

Blandhol, C., Bonney, J., Mogstad, M., and Torgovitsky, A. (2022). When is TSLS actually LATE? Technical report, National Bureau of Economic Research.

Brazilian Institute of Geography and Statistics (2012). Brazil demographic census 2010. Technical report. Rio de Janeiro, Brazil, 2012.

Carneiro, P., Heckman, J. J., and Vytlacil, E. (2011). Estimating marginal returns to education. *American Economic Review*, 101(6):2754–2781.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica*, 71(5):1405–1441.

Cole, H., Court, P., Deutschmann, J. W., and Lipscomb, M. (2025). Identifying the costs of sewer and stormwater blockages. Seminar presentation. Cape Town.

Coury, M., Kitagawa, T., Shertzer, A., and Turner, M. (2022). The value of piped water and sewers: Evidence from 19th century Chicago. Technical report, National Bureau of Economic Research.

Department of Statistics (Jordan) (2015). Jordan population and housing census 2015. Technical report. Department of Statistics, Aman Jordan.

Ferrie, J. P. and Troesken, W. (2008). Water and Chicago's mortality transition, 1850–1925. *Explorations in Economic History*, 45(1):1–16.

Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., Haller, L., and Colford Jr, J. M. (2005). Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *The Lancet infectious diseases*, 5(1):42–52.

Galiani, S., Gertler, P., and Schargrodsky, E. (2005). Water for life: The impact of the privatization of water services on child mortality. *Journal of Political Economy*, 113(1):83–120.

Gamper-Rabindran, S., Khan, S., and Timmins, C. (2010). The impact of piped water provision on infant mortality in Brazil: A quantile panel data approach. *Journal of Development Economics*, 92(2):188–200.

Gendron-Carrier, N., Gonzalez-Navarro, M., Polloni, S., and Turner, M. A. (2022). Subways and urban air pollution. *American Economic Journal: Applied Economics*, 14(1):164–96.

Ghani, E., Goswami, A. G., and Kerr, W. R. (2016). Highway to success: The impact of the golden quadrilateral project for the location and performance of Indian manufacturing. *The Economic Journal*, 126(591):317–357.

Gibson, J., McKenzie, D., and Rohorua, H. (2014). Development impacts of seasonal and temporary migration: A review of evidence from the Pacific and Southeast Asia. *Asia & the Pacific Policy Studies*, 1(1):18–32.

Heblich, S., Redding, S. J., and Sturm, D. M. (2020). The making of the modern metropolis: evidence from London. *The Quarterly Journal of Economics*, 135(4):2059–2133.

Henderson, J. V. and Turner, M. A. (2020). Urbanization in the developing world: too early or too slow? *Journal of Economic Perspectives*, 34(3):150–73.

Hutton, G., Haller, L., and Bartram, J. (2007). Economic and health effects of increasing coverage of low cost household drinking water supply and sanitation interventions. *World Health Organization*.

Imbens, G. W. and Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512.

Jedwab, R. and Storeygard, A. (2022). The average and heterogeneous effects of transportation investments: Evidence from sub-saharan africa 1960–2010. *Journal of the European Economic Association*, 20(1):1–38.

Kesztenbaum, L. and Rosenthal, J. L. (2017). Sewers' diffusion and the decline of mortality: The case of Paris, 1880–1914. *Journal of Urban Economics*, 98:174–186.

Kolesár, M. and Plagborg-Møller, M. (2024). Dynamic causal effects in a nonlinear world: the good, the bad, and the ugly. *arXiv preprint arXiv:2411.10415*.

Lagakos, D., Marshall, S., Mobarak, A. M., Vernot, C., and Waugh, M. E. (2020). Migration costs and observational returns to migration in the developing world. *Journal of Monetary Economics*, 113:138–154.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, 95:391–413.

Lipscomb, M., Mobarak, A. M., and Barham, T. (2013). Development effects of electrification: Evidence from the topographic placement of hydropower plants in Brazil. *American Economic Journal: Applied Economics*, 5(2):200–231.

Mara, D. (1996). *Low-cost sewerage*. John Wiley London.

NASA JPL (2013). Nasa shuttle radar topography mission global 1 arc second [data set]. Technical report. Last accessed 19 May 2022.

NASA/METI/AIST/Japan Spacesystems and US/Japan ASTER Science Team (2019). ASTER global digital elevation model v003. Technical report. Last accessed 19 May 2022.

National Administrative Department of Statistics (2018). Colombia population and housing census 2018. Technical report. Bogotá, Colombia: National Administrative Department of Statistics.

National Bureau of Statistics (Tanzania), Office of the Chief Government Statistician (2012). Tanzania population and housing census 2012. Technical report. National Bureau of Statistics (Tanzania), Office of the Chief Government Statistician (Zanzibar).

Ogasawara, K. and Matsushita, Y. (2018). Public health and multiple-phase mortality decline: Evidence from industrializing Japan. *Economics & Human Biology*, 29:198–210.

Selendy, J. M. (2011). *Water and sanitation-related diseases and the environment: challenges, interventions, and preventive measures*. John Wiley & Sons.

Słoczyński, T. (2021). When should we (not) interpret linear IV estimands as LATE? *unpublished manuscript*.

Statistics South Africa (2011). Census 2011. Technical report. Pretoria, South Africa: Statistics South Africa.

UN DESA Population Division (2018). World urbanization prospects: the 2018 revision.

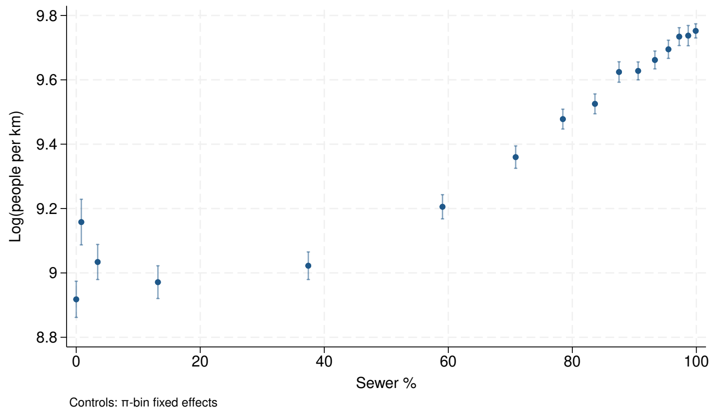
Tsivanidis, N. (2019). Evaluating the impact of urban transit infrastructure: Evidence from Bogota's Transmilenio.

Uuemaa, E., Ahi, S., Montibeller, B., Muru, M., and Kmoch, A. (2020). Vertical accuracy of freely available global digital elevation models (ASTER, AW3D30, MERIT, TanDEM-X, SRTM, and NASADEM). *Remote Sensing*, 12(21):3482.

Zárate, R. D. (2022). Spatial misallocation, informality, and transit improvements: Evidence from Mexico City.

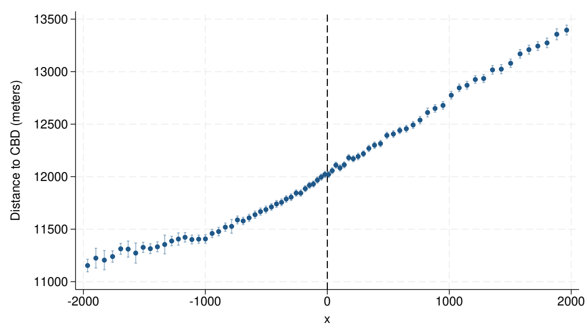
Appendix A Supplemental results

Figure A1: Logarithm of tract population density vs sewer share



Note: *Mean log population by tract sewer share. Log tract population density increases from about 8.8 to about 9.8 as sewer share increases from zero to one. On average, each 1 percentage point increase in sewer share is associated with about a 1% increase in population density. Figure based on the estimation sample described in column 2 of table 1.*

Figure A2: A placebo test



Note: *A binscatter plot with horizontal distance from the central basin divide on the x-axis. y-axis shows bin mean of distance to* CBD *net of radial-bin mean. That this plot is continuous reassures us that we have not introduced an unintended sampling restriction. Figure based on the estimation sample described in column 2 of table 1.*

Table A1: Estimation sample by country

	(1) Cities	(2) π-bins	(3) Tracts	(4) Share inside	(5) Tract area	(6) Pop. Density	(7) Sewer share
Brazil	56	1,155	22,810	0.54	0.32	16,792	0.67
Colombia	18	355	23,199	0.53	0.10	28,880	0.72
Jordan	2	8	26	0.35	2.35	878	0.43
South Africa	12	273	3,177	0.52	0.47	8,358	0.77
Tanzania	7	105	1,593	0.65	0.13	24,936	0.08

Note: *Census data for Jordan reports households, not people. Columns (4-7) are tract weighted averages. Population density is people per* KM^2 *and tract area is* KM^2.

Table A2: Sewers and log tract population density, clustered errors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.8562***	0.8562***	0.8562***	0.8108***	0.8108***	0.8108***	0.7398***	0.7398***	0.7398***
	(0.0538)	(0.0538)	(0.0538)	(0.0540)	(0.0540)	(0.0540)	(0.0553)	(0.0553)	(0.0553)
2. First-stage									
Outside	-0.00800	-0.00554	-0.00509	-0.01404	-0.00335	-0.00448	-0.00984	-0.00228	-0.00461
	(0.008763)	(0.010878)	(0.010455)	(0.009747)	(0.010195)	(0.009884)	(0.009193)	(0.011392)	(0.010909)
x*Outside	-0.00005**		-0.00005**	-0.00006***		-0.00006***	-0.00007***		-0.00007***
	(0.000025)		(0.000024)	(0.000017)		(0.000017)	(0.000014)		(0.000014)
ΔElev*Outside		-0.00018	-0.00013		-0.00045**	-0.00040*		-0.00025	-0.00021
		(0.000335)	(0.000318)		(0.000218)	(0.000212)		(0.000223)	(0.000230)
3. IV									
Sewer share	1.8284**	3.5182	2.0739**	3.9819***	2.5200*	4.1937***	6.0387***	0.4847	5.9413***
	(0.8031)	(3.7855)	(0.8836)	(1.0333)	(1.4999)	(1.0471)	(1.3402)	(3.0922)	(1.3008)
N	50805	50805	50805	50805	50805	50805	50805	50805	50805
F	95.92	7.335	65.13	83.40	20.71	62.33	76.21	6.163	52.35
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Standard errors clustered by π-bin in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Table A3: Sewers and log tract population density, Brazil

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.7815***	0.7815***	0.7815***	0.7633***	0.7633***	0.7633***	0.6971***	0.6971***	0.6971***
	(0.0382)	(0.0382)	(0.0382)	(0.0382)	(0.0382)	(0.0382)	(0.0386)	(0.0386)	(0.0386)
2. First-stage									
Outside	-0.00340	-0.01286**	-0.01260**	-0.00934*	-0.00914	-0.01031	-0.01078*	-0.01141*	-0.01303*
	(0.005409)	(0.005710)	(0.005710)	(0.005609)	(0.006390)	(0.006390)	(0.005761)	(0.006783)	(0.006775)
x*Outside	-0.00002***		-0.00002***	-0.00003***		-0.00003***	-0.00006***		-0.00006***
	(0.000006)		(0.000006)	(0.000008)		(0.000008)	(0.000010)		(0.000010)
ΔElev*Outside		0.00047***	0.00048***		-0.00000	0.00005		0.00003	0.00011
		(0.000078)	(0.000079)		(0.000125)	(0.000125)		(0.000137)	(0.000138)
3. IV									
Sewer share	3.9580**	-0.6409	0.6877	7.4377***	2.5457	7.2668***	6.8251***	4.0942	6.6577***
	(1.6653)	(0.9821)	(0.7433)	(1.7599)	(2.7991)	(1.7245)	(1.1310)	(2.9249)	(1.1064)
4. SATE									
Sewered	3.8242**	-0.5164	0.7064	6.6993***	3.7359	6.5523***	5.7247***	5.6278**	5.5581***
	(1.1973)	(0.9259)	(0.7211)	(0.7032)	(2.3482)	(0.7059)	(0.5455)	(1.9707)	(0.5464)
N	22810	22810	22810	22810	22810	22810	22810	22810	22810
F	5.354	13.53	12.83	11.87	1.364	7.953	25.74	1.854	17.35
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Table A4: Sewers and log tract population density, Colombia

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	1.1226***	1.1226***	1.1226***	1.0581***	1.0581***	1.0581***	0.9739***	0.9739***	0.9739***
	(0.0373)	(0.0373)	(0.0373)	(0.0398)	(0.0398)	(0.0398)	(0.0391)	(0.0391)	(0.0391)
2. First-stage									
Outside	-0.00379	0.03140***	0.02963***	-0.01091**	0.01797***	0.01859***	-0.00464	0.02134***	0.01780**
	(0.005131)	(0.007001)	(0.006991)	(0.005155)	(0.006821)	(0.006812)	(0.005384)	(0.007284)	(0.007231)
x*Outside	-0.00010***		-0.00009***	-0.00010***		-0.00010***	-0.00010***		-0.00010***
	(0.000006)		(0.000006)	(0.000007)		(0.000007)	(0.000008)		(0.000008)
ΔElev*Outside		-0.00142***	-0.00123***		-0.00108***	-0.00104***		-0.00080***	-0.00078***
		(0.000190)	(0.000190)		(0.000175)	(0.000176)		(0.000191)	(0.000188)
3. IV									
Sewer share	1.3582***	4.2612***	2.4230***	2.4031***	3.9623***	3.0252***	4.3853***	2.7042*	4.1823***
	(0.3131)	(0.5882)	(0.3066)	(0.3828)	(0.8998)	(0.3714)	(0.5346)	(1.4053)	(0.5019)
4. SATE									
Sewered	2.5266***	4.6208***	3.5564***	2.6972***	4.2664***	3.3492***	4.6793***	4.5623**	4.7469***
	(0.4954)	(0.5323)	(0.3999)	(0.3952)	(0.8981)	(0.3942)	(0.4959)	(1.4116)	(0.4898)
N	23199	23199	23199	23199	23199	23199	23199	23199	23199
F	141.2	78.87	133.2	113.6	35.52	96.18	72.91	12.48	56.72
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Table A5: Sewers and log tract population density, South Africa

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	-0.0882	-0.0882	-0.0882	-0.0152	-0.0152	-0.0152	-0.2025*	-0.2025*	-0.2025*
	(0.0974)	(0.0974)	(0.0974)	(0.0997)	(0.0997)	(0.0997)	(0.1053)	(0.1053)	(0.1053)
2. First-stage									
Outside	-0.06302***	-0.06381***	-0.06381***	-0.07186***	-0.05879***	-0.05852***	-0.04186**	-0.03964**	-0.03786*
	(0.016134)	(0.016741)	(0.016731)	(0.016150)	(0.017462)	(0.017434)	(0.016607)	(0.019484)	(0.019398)
x*Outside	-0.00000		0.00000	0.00001		0.00001	0.00007**		0.00007**
	(0.000019)		(0.000019)	(0.000022)		(0.000022)	(0.000029)		(0.000029)
ΔElev*Outside		0.00005	0.00005		-0.00083*	-0.00082*		-0.00038	-0.00023
		(0.000289)	(0.000288)		(0.000440)	(0.000439)		(0.000514)	(0.000508)
3. IV									
Sewer share	0.3373	0.4492	0.4491	-0.3493	-1.2357	-1.4434	-3.1915**	-1.9157	-3.7750**
	(1.1152)	(1.1193)	(1.1192)	(0.9850)	(0.9445)	(0.9506)	(1.5008)	(1.7553)	(1.5823)
4. SATE									
Sewered	0.5615	0.8808	0.8111	0.1251	-0.6076	-0.8920	0.2899	0.8870	-0.3167
	(0.9760)	(0.9841)	(0.9754)	(0.8668)	(0.8497)	(0.8281)	(1.3760)	(1.5434)	(1.3602)
N	3177	3177	3177	3177	3177	3177	3177	3177	3177
F	7.539	7.548	5.030	9.938	11.49	7.764	7.301	4.122	4.939
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Table A6: Sewers and log tract population density, CBD distance control

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.8490***	0.8490***	0.8490***	0.8116***	0.8116***	0.8116***	0.7400***	0.7400***	0.7400***
	(0.0260)	(0.0260)	(0.0260)	(0.0266)	(0.0266)	(0.0266)	(0.0266)	(0.0266)	(0.0266)
2. First-stage									
Outside	-0.00433	0.00320	0.00319	-0.01139***	0.00460	0.00327	-0.00790**	0.00622	0.00381
	(0.003552)	(0.003977)	(0.003973)	(0.003645)	(0.004399)	(0.004389)	(0.003772)	(0.004739)	(0.004717)
x*Outside	-0.00005***		-0.00004***	-0.00005***		-0.00005***	-0.00007***		-0.00006***
	(0.000004)		(0.000004)	(0.000005)		(0.000005)	(0.000006)		(0.000006)
ΔElev*Outside		-0.00037***	-0.00032***		-0.00066***	-0.00061***		-0.00052***	-0.00047***
		(0.000080)	(0.000079)		(0.000100)	(0.000099)		(0.000112)	(0.000110)
3. IV									
Sewer share	1.9919***	5.2171***	2.7225***	4.5107***	2.9927***	4.4164***	6.4736***	1.3674	5.7667***
	(0.4328)	(1.1667)	(0.4299)	(0.5569)	(0.7517)	(0.4845)	(0.6864)	(1.0500)	(0.5872)
N	50805	50805	50805	50805	50805	50805	50805	50805	50805
F	64.46	16.22	50.42	63.10	31.55	57.06	64.42	14.70	50.19
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Table A7: Sewers and log tract population density, universal piped water sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.7262***	0.7262***	0.7262***	0.6807***	0.6807***	0.6807***	0.6470***	0.6470***	0.6470***
	(0.0402)	(0.0402)	(0.0402)	(0.0405)	(0.0405)	(0.0405)	(0.0407)	(0.0407)	(0.0407)
2. First-stage									
Outside	0.00328	0.00805*	0.00816*	0.00823**	0.01383***	0.01349***	0.01173***	0.00999*	0.00791
	(0.003794)	(0.004385)	(0.004386)	(0.004063)	(0.005122)	(0.005123)	(0.004210)	(0.005438)	(0.005450)
x*Outside	-0.00002***		-0.00001***	-0.00002***		-0.00002***	-0.00005***		-0.00005***
	(0.000005)		(0.000005)	(0.000006)		(0.000006)	(0.000006)		(0.000007)
ΔElev*Outside		-0.00020***	-0.00018**		-0.00020**	-0.00019*		0.00012	0.00014
		(0.000075)	(0.000075)		(0.000104)	(0.000104)		(0.000123)	(0.000123)
3. IV									
Sewer share	2.8401*	12.3844**	5.5065***	5.7613***	13.5257***	6.8550***	5.1458***	7.2149***	4.9844***
	(1.5991)	(4.8353)	(1.6815)	(1.7905)	(5.0659)	(1.8122)	(0.9430)	(2.5349)	(0.9287)
N	26801	26801	26801	26801	26801	26801	26801	26801	26801
F	6.054	3.477	5.871	9.000	3.926	7.057	28.81	5.783	19.65
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Table A8: Sewers and log tract population density, piped water control

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.6701***	0.6701***	0.6701***	0.6535***	0.6535***	0.6535***	0.6191***	0.6191***	0.6191***
	(0.0253)	(0.0253)	(0.0253)	(0.0257)	(0.0257)	(0.0257)	(0.0261)	(0.0261)	(0.0261)
2. First-stage									
Outside	-0.00852**	-0.00797**	-0.00759**	-0.01254***	-0.00604	-0.00679	-0.01028***	-0.00621	-0.00806*
	(0.003344)	(0.003702)	(0.003701)	(0.003484)	(0.004156)	(0.004158)	(0.003614)	(0.004504)	(0.004509)
x*Outside	-0.00004***		-0.00004***	-0.00004***		-0.00004***	-0.00006***		-0.00006***
	(0.000004)		(0.000004)	(0.000005)		(0.000005)	(0.000006)		(0.000006)
ΔElev*Outside		-0.00008	-0.00004		-0.00027***	-0.00024***		-0.00012	-0.00009
		(0.000065)	(0.000065)		(0.000086)	(0.000087)		(0.000098)	(0.000099)
3. IV									
Sewer share	1.6785***	0.4453	1.7903***	4.3971***	1.5131	4.7165***	6.8254***	-1.0872	6.8008***
	(0.4190)	(1.4566)	(0.4213)	(0.6794)	(1.1153)	(0.6736)	(0.8078)	(1.8273)	(0.7994)
N	50805	50805	50805	50805	50805	50805	50805	50805	50805
F	73.11	5.180	48.87	44.59	11.86	32.32	53.15	4.353	35.72
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Table A9: Sewers and log tract population density, universal garbage collection sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.6808***	0.6808***	0.6808***	0.6546***	0.6546***	0.6546***	0.6159***	0.6159***	0.6159***
	(0.0303)	(0.0303)	(0.0303)	(0.0309)	(0.0309)	(0.0309)	(0.0319)	(0.0319)	(0.0319)
2. First-stage									
Outside	-0.00595	-0.00681	-0.00635	-0.00345	0.00045	-0.00044	0.00233	0.00006	-0.00291
	(0.003903)	(0.004464)	(0.004465)	(0.004124)	(0.005189)	(0.005204)	(0.004308)	(0.005576)	(0.005593)
x*Outside	-0.00004***		-0.00004***	-0.00005***		-0.00005***	-0.00006***		-0.00007***
	(0.000005)		(0.000005)	(0.000006)		(0.000006)	(0.000007)		(0.000007)
ΔElev*Outside		-0.00002	0.00002		-0.00017	-0.00012		0.00014	0.00020
		(0.000076)	(0.000077)		(0.000108)	(0.000109)		(0.000128)	(0.000129)
3. IV									
Sewer share	0.4120	-5.6432	0.3229	2.3111***	6.9606*	2.6003***	4.2347***	0.3261	3.9387***
	(0.5277)	(3.9295)	(0.5287)	(0.5488)	(4.0271)	(0.5560)	(0.6116)	(3.4207)	(0.5894)
N	32193	32193	32193	32193	32193	32193	32193	32193	32193
F	35.07	1.810	23.39	40.76	1.681	27.55	48.71	1.146	33.43
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.*

Table A10: Sewers and log tract population density, garbage collection control

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.6474***	0.6474***	0.6474***	0.6294***	0.6294***	0.6294***	0.6027***	0.6027***	0.6027***
	(0.0250)	(0.0250)	(0.0250)	(0.0255)	(0.0255)	(0.0255)	(0.0260)	(0.0260)	(0.0260)
2. First-stage									
Outside	-0.00863***	-0.00653*	-0.00617*	-0.01338***	-0.00488	-0.00572	-0.00876**	-0.00219	-0.00408
	(0.003339)	(0.003683)	(0.003682)	(0.003457)	(0.004154)	(0.004158)	(0.003587)	(0.004499)	(0.004507)
x*Outside	-0.00004***		-0.00004***	-0.00004***		-0.00004***	-0.00006***		-0.00006***
	(0.000004)		(0.000004)	(0.000005)		(0.000005)	(0.000006)		(0.000006)
ΔElev*Outside		-0.00015**	-0.00011		-0.00036***	-0.00032***		-0.00022**	-0.00019*
		(0.000064)	(0.000065)		(0.000087)	(0.000088)		(0.000101)	(0.000102)
3. IV									
Sewer share	1.5568***	2.6558*	1.8795***	4.1622***	2.0731**	4.4617***	6.8735***	0.2707	6.7309***
	(0.4326)	(1.4076)	(0.4380)	(0.6031)	(0.9821)	(0.5883)	(0.8113)	(1.7161)	(0.7804)
N	50773	50773	50773	50773	50773	50773	50773	50773	50773
F	66.75	6.980	45.41	54.31	16.41	40.82	53.14	5.181	36.74
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.*

Table A11: Sewers and log tract population density, dropping hilly areas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.8040***	0.8040***	0.8040***	0.7622***	0.7622***	0.7622***	0.7017***	0.7017***	0.7017***
	(0.0258)	(0.0258)	(0.0258)	(0.0270)	(0.0270)	(0.0270)	(0.0270)	(0.0270)	(0.0270)
2. First-stage									
Outside	-0.00813**	0.00511	0.00442	-0.01091***	-0.00098	-0.00214	-0.00780**	-0.00466	-0.00663
	(0.003757)	(0.004628)	(0.004619)	(0.003856)	(0.005066)	(0.005059)	(0.003980)	(0.005397)	(0.005397)
x*Outside	-0.00006***		-0.00006***	-0.00005***		-0.00005***	-0.00007***		-0.00007***
	(0.000005)		(0.000005)	(0.000006)		(0.000006)	(0.000007)		(0.000007)
ΔElev*Outside		-0.00067***	-0.00058***		-0.00046***	-0.00041***		-0.00007	-0.00005
		(0.000135)	(0.000135)		(0.000154)	(0.000153)		(0.000167)	(0.000166)
3. IV									
Sewer share	1.2375***	2.6274***	1.6361***	3.3462***	1.4703	3.5593***	5.2988***	-5.4821	5.3133***
	(0.3332)	(0.7219)	(0.3128)	(0.4992)	(1.0229)	(0.4879)	(0.6141)	(4.5840)	(0.6136)
N	45081	45081	45081	45081	45081	45081	45081	45081	45081
F	90.36	19.91	68.38	58.77	9.472	42.14	61.82	1.472	41.26
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses.* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Sewers and log tract population density, dropping areas near municipal boundaries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.8506***	0.8506***	0.8506***	0.8012***	0.8012***	0.8012***	0.7325***	0.7325***	0.7325***
	(0.0266)	(0.0266)	(0.0266)	(0.0277)	(0.0277)	(0.0277)	(0.0275)	(0.0275)	(0.0275)
2. First-stage									
Outside	-0.01114***	-0.00915**	-0.00862**	-0.01533***	-0.00810*	-0.00875*	-0.01566***	-0.00894*	-0.01083**
	(0.003662)	(0.004036)	(0.004034)	(0.003758)	(0.004475)	(0.004474)	(0.003873)	(0.004819)	(0.004825)
x*Outside	-0.00006***		-0.00006***	-0.00004***		-0.00004***	-0.00006***		-0.00006***
	(0.000004)		(0.000004)	(0.000005)		(0.000005)	(0.000006)		(0.000006)
ΔElev*Outside		-0.00017**	-0.00011		-0.00030***	-0.00028***		-0.00023**	-0.00020*
		(0.000077)	(0.000078)		(0.000101)	(0.000101)		(0.000113)	(0.000113)
3. IV									
Sewer share	1.3815***	1.3712	1.5784***	3.5948***	1.3122	4.0367***	4.9751***	-0.3898	4.9051***
	(0.3466)	(1.1136)	(0.3481)	(0.6213)	(1.0036)	(0.6168)	(0.5652)	(1.1661)	(0.5504)
N	48009	48009	48009	48009	48009	48009	48009	48009	48009
F	95.96	9.122	64.83	41.90	13.91	31.00	64.74	9.891	44.39
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses.* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: Sewers and electricity access

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.0857***	0.0857***	0.0857***	0.0869***	0.0869***	0.0869***	0.0837***	0.0837***	0.0837***
	(0.0033)	(0.0033)	(0.0033)	(0.0035)	(0.0035)	(0.0035)	(0.0036)	(0.0036)	(0.0036)
2. First-stage									
Outside	-0.00798**	-0.00557	-0.00512	-0.01403***	-0.00335	-0.00447	-0.00978***	-0.00222	-0.00456
	(0.003561)	(0.003946)	(0.003946)	(0.003647)	(0.004385)	(0.004386)	(0.003772)	(0.004707)	(0.004715)
x*Outside	-0.00005***		-0.00005***	-0.00006***		-0.00006***	-0.00007***		-0.00007***
	(0.000004)		(0.000004)	(0.000005)		(0.000005)	(0.000006)		(0.000006)
ΔElev*Outside		-0.00018**	-0.00012		-0.00045***	-0.00040***		-0.00025**	-0.00021*
		(0.000076)	(0.000077)		(0.000099)	(0.000099)		(0.000110)	(0.000110)
3. IV									
Sewer share	0.0636**	0.2252**	0.0615**	0.0942***	0.1681***	0.0881***	0.0404	0.1283	0.0357
	(0.0269)	(0.0971)	(0.0264)	(0.0281)	(0.0611)	(0.0253)	(0.0285)	(0.1189)	(0.0273)
4. SATE									
Sewered	0.1065***	0.4025***	0.1094***	0.0858**	0.2265***	0.0876***	0.0328	0.1904	0.0310
	(0.0288)	(0.0876)	(0.0287)	(0.0271)	(0.0567)	(0.0250)	(0.0273)	(0.1083)	(0.0264)
N	50773	50773	50773	50773	50773	50773	50773	50773	50773
F	92.58	6.983	62.82	81.02	20.14	60.56	73.19	5.881	50.28
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Table A14: Sewers and literacy rate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.0005	0.0005	0.0005	-0.0042	-0.0042	-0.0042	0.0229***	0.0229***	0.0229***
	(0.0180)	(0.0180)	(0.0180)	(0.0182)	(0.0182)	(0.0182)	(0.0032)	(0.0032)	(0.0032)
2. First-stage									
Outside	-0.00807**	-0.00567	-0.00523	-0.01411***	-0.00342	-0.00454	-0.00987***	-0.00230	-0.00463
	(0.003562)	(0.003947)	(0.003946)	(0.003648)	(0.004386)	(0.004387)	(0.003773)	(0.004708)	(0.004716)
x*Outside	-0.00005***		-0.00005***	-0.00006***		-0.00006***	-0.00007***		-0.00007***
	(0.000004)		(0.000004)	(0.000005)		(0.000005)	(0.000006)		(0.000006)
ΔElev*Outside		-0.00018**	-0.00012		-0.00045***	-0.00040***		-0.00026**	-0.00021*
		(0.000076)	(0.000077)		(0.000099)	(0.000099)		(0.000110)	(0.000110)
3. IV									
Sewer share	0.1230**	-0.4301**	0.1107*	0.0457	0.0764	0.0676**	0.0288	0.0345	0.0377
	(0.0603)	(0.1979)	(0.0598)	(0.0398)	(0.0821)	(0.0343)	(0.0333)	(0.0874)	(0.0320)
4. SATE									
Sewered	0.1669**	-0.2995*	0.1620*	0.0442	0.0246	0.0596	0.0339	-0.0036	0.0408
	(0.0628)	(0.1470)	(0.0633)	(0.0375)	(0.0698)	(0.0330)	(0.0315)	(0.0911)	(0.0304)
N	50777	50777	50777	50777	50777	50777	50777	50777	50777
F	92.19	7.022	62.55	80.44	20.25	60.19	72.73	5.949	49.98
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Table A15: Sewers and log mean tract income, Brazil and South Africa only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.4599***	0.4599***	0.4599***	0.4153***	0.4153***	0.4153***	0.4030***	0.4030***	0.4030***
	(0.0141)	(0.0141)	(0.0141)	(0.0138)	(0.0138)	(0.0138)	(0.0146)	(0.0146)	(0.0146)
2. First-stage									
Outside	-0.01030**	-0.01912***	-0.01893***	-0.01667***	-0.01530**	-0.01620***	-0.01501***	-0.01480**	-0.01607**
	(0.005138)	(0.005410)	(0.005410)	(0.005300)	(0.006000)	(0.005999)	(0.005454)	(0.006410)	(0.006407)
x*Outside	-0.00002***		-0.00002***	-0.00003***		-0.00003***	-0.00005***		-0.00005***
	(0.000006)		(0.000006)	(0.000008)		(0.000008)	(0.000010)		(0.000010)
ΔElev*Outside		0.00045***	0.00046***		-0.00006	-0.00002		0.00001	0.00005
		(0.000076)	(0.000076)		(0.000121)	(0.000121)		(0.000133)	(0.000134)
3. IV									
Sewer share	2.8360***	1.7680***	2.6436***	1.6207***	-0.9938	1.6202***	0.8452**	-1.9792*	0.8242**
	(0.8702)	(0.5080)	(0.4942)	(0.4915)	(0.7854)	(0.4917)	(0.3719)	(1.1642)	(0.3689)
4. SATE									
Sewered	2.9623***	1.7721***	2.6010***	1.5356***	-1.0600	1.5340***	0.6886*	-1.7475*	0.6654
	(0.4819)	(0.4382)	(0.3427)	(0.3589)	(0.6594)	(0.3602)	(0.3501)	(0.7790)	(0.3482)
N	25986	25986	25986	25986	25986	25986	25986	25986	25986
F	6.746	14.92	13.36	12.34	4.947	8.238	19.05	3.788	12.74
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Table A16: Sewers and apartments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.1069***	0.1069***	0.1069***	0.0918***	0.0918***	0.0918***	0.0850***	0.0850***	0.0850***
	(0.0050)	(0.0050)	(0.0050)	(0.0052)	(0.0052)	(0.0052)	(0.0055)	(0.0055)	(0.0055)
2. First-stage									
Outside	-0.00786**	-0.00564	-0.00533	-0.01451***	-0.00379	-0.00490	-0.01017***	-0.00272	-0.00507
	(0.003634)	(0.004010)	(0.004009)	(0.003715)	(0.004451)	(0.004452)	(0.003837)	(0.004763)	(0.004771)
x*Outside	-0.00005***		-0.00005***	-0.00006***		-0.00006***	-0.00007***		-0.00007***
	(0.000004)		(0.000004)	(0.000005)		(0.000005)	(0.000006)		(0.000006)
ΔElev*Outside		-0.00016**	-0.00011		-0.00046***	-0.00040***		-0.00025**	-0.00021*
		(0.000077)	(0.000078)		(0.000101)	(0.000101)		(0.000111)	(0.000112)
3. IV									
Sewer share	0.4902***	-1.1238**	0.4596***	0.5533***	0.2352	0.5374***	0.3395***	0.4501	0.3700***
	(0.0855)	(0.5243)	(0.0847)	(0.0925)	(0.1796)	(0.0886)	(0.0971)	(0.3731)	(0.0975)
4. SATE									
Sewered	0.4653***	-1.1921**	0.4437***	0.5612***	0.1183	0.5248***	0.3656***	0.2528	0.3840***
	(0.0753)	(0.3770)	(0.0756)	(0.0791)	(0.1783)	(0.0772)	(0.0902)	(0.3492)	(0.0899)
N	49182	49182	49182	49182	49182	49182	49182	49182	49182
F	89.93	6.059	60.80	81.42	20.38	60.73	73.32	5.857	50.27
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

Table A17: Sewers and garbage collection

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.2127***	0.2127***	0.2127***	0.2006***	0.2006***	0.2006***	0.1980***	0.1980***	0.1980***
	(0.0050)	(0.0050)	(0.0050)	(0.0052)	(0.0052)	(0.0052)	(0.0055)	(0.0055)	(0.0055)
2. First-stage									
Outside	-0.00798**	-0.00557	-0.00512	-0.01403***	-0.00335	-0.00447	-0.00978***	-0.00222	-0.00456
	(0.003561)	(0.003946)	(0.003946)	(0.003647)	(0.004385)	(0.004386)	(0.003772)	(0.004707)	(0.004715)
x*Outside	-0.00005***		-0.00005***	-0.00006***		-0.00006***	-0.00007***		-0.00007***
	(0.000004)		(0.000004)	(0.000005)		(0.000005)	(0.000006)		(0.000006)
ΔElev*Outside		-0.00018**	-0.00012		-0.00045***	-0.00040***		-0.00025**	-0.00021*
		(0.000076)	(0.000077)		(0.000099)	(0.000099)		(0.000110)	(0.000110)
3. IV									
Sewer share	0.3919***	0.1196	0.3899***	0.4750***	0.3071***	0.4698***	0.4042***	0.2270	0.3989***
	(0.0484)	(0.1882)	(0.0479)	(0.0547)	(0.1104)	(0.0510)	(0.0558)	(0.2024)	(0.0546)
4. SATE									
Sewered	0.5440***	0.7084***	0.5671***	0.4436***	0.4287***	0.4582***	0.3836***	0.4265*	0.3855***
	(0.0488)	(0.1854)	(0.0493)	(0.0450)	(0.1100)	(0.0440)	(0.0488)	(0.1855)	(0.0482)
N	50773	50773	50773	50773	50773	50773	50773	50773	50773
F	92.58	6.983	62.82	81.02	20.14	60.56	73.19	5.881	50.28
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * $p <$ 0.10, ** $p < 0.05$, *** $p < 0.01$.*

Table A18: Sewers and piped water access

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. OLS									
Sewer share	0.2544***	0.2544***	0.2544***	0.2263***	0.2263***	0.2263***	0.2189***	0.2189***	0.2189***
	(0.0055)	(0.0055)	(0.0055)	(0.0055)	(0.0055)	(0.0055)	(0.0058)	(0.0058)	(0.0058)
2. First-stage									
Outside	-0.00800**	-0.00554	-0.00509	-0.01404***	-0.00335	-0.00448	-0.00984***	-0.00228	-0.00461
	(0.003561)	(0.003946)	(0.003946)	(0.003648)	(0.004385)	(0.004386)	(0.003773)	(0.004708)	(0.004716)
x*Outside	-0.00005***		-0.00005***	-0.00006***		-0.00006***	-0.00007***		-0.00007***
	(0.000004)		(0.000004)	(0.000005)		(0.000005)	(0.000006)		(0.000006)
ΔElev*Outside		-0.00018**	-0.00013		-0.00045***	-0.00040***		-0.00025**	-0.00021*
		(0.000076)	(0.000077)		(0.000099)	(0.000099)		(0.000110)	(0.000110)
3. IV									
Sewer share	0.3605***	0.5362***	0.3812***	0.7172***	0.6817***	0.7443***	0.4622***	0.5831***	0.4903***
	(0.0552)	(0.1897)	(0.0547)	(0.0700)	(0.1246)	(0.0663)	(0.0640)	(0.2256)	(0.0641)
4. SATE									
Sewered	0.5239***	1.0759***	0.5646***	0.6950***	0.8764***	0.7465***	0.4343***	0.9067***	0.4723***
	(0.0607)	(0.1629)	(0.0610)	(0.0495)	(0.1021)	(0.0480)	(0.0545)	(0.1736)	(0.0541)
N	50805	50805	50805	50805	50805	50805	50805	50805	50805
F	92.34	7.061	62.70	81.16	20.15	60.65	73.25	5.924	50.32
Elevation	Y	Y	Y	Y	Y	Y	Y	Y	Y
π-bins	Y	Y	Y	Y	Y	Y	Y	Y	Y
x	Y	Y	Y						
seg$\times$x				Y	Y	Y			
π-bins$\times$x							Y	Y	Y

Note: *Sample is described by column 2 of table 1. Robust standard errors in parentheses. * $p <$ 0.10, ** $p < 0.05$, *** $p < 0.01$.*

Appendix B Data

A. Cities

The UN DESA World Urbanization Prospects data (UN DESA Population Division (2018)) is a census of all cities that had a population 300,000 or more in 2014. These data report coordinates for the city center – we frequently refer to this point as the *central business district*. We focus attention on areas that are both within 75KM of the city center and near the boundary of the drainage basin containing the city center.

Intersecting with the census data discussed below, we estimate treatment effects using all cities in the UN Cities data in Brazil, Colombia, South Africa, Jordan, and Tanzania. We can also evaluate counterfactuals in these cities.

B. Sewers

Each of the countries we study provide comprehensive surveys of sewer access at granular geographies in the 2010s. We calculate the share of households in each census geography with sewer access and map the extent of tracts with sewers.

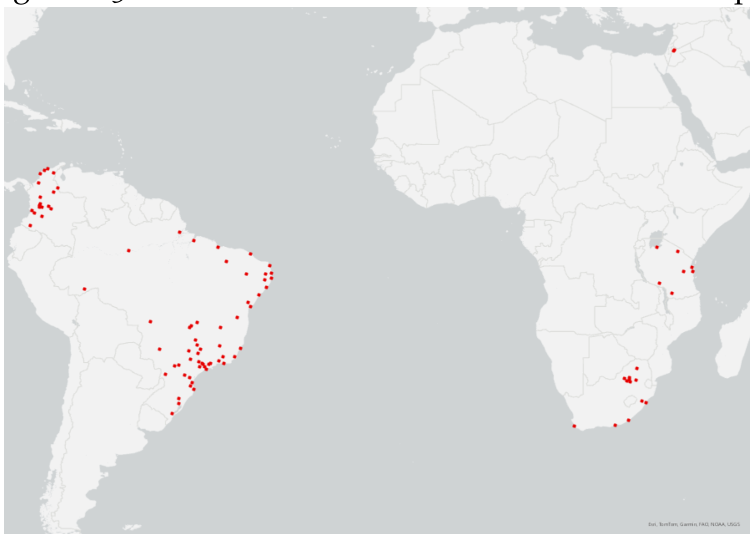
The census in Brazil (Brazilian Institute of Geography and Statistics, 2012) asks households, "Is the bathroom or toilet drain connected to the public sewer system?" The results are reported at the *setores* (English: *sectors*) geographic unit with counts of households affirming and total number of households in the *setor*.

The Colombian census (National Administrative Department of Statistics, 2018) reports counts of households indicating sewer service in response to, "Does your house have sewage service?" This is released at different details of granularity: in urban areas, this is available at the *manzana*-level (English: *block, square*) correseponding to a very fine spatial detail; in rural areas, this is available at the *seccion* (English: *section*) which is larger.

The South African census (Statistics South Africa, 2011) counts households with sewer access using this question: "Is the main type of toilet facility used by this household a flush toilet connected to sewerage system?" The results are reported at a geography the *Small Area Layer* geography, which is between an "enumeration area" and "sub-place."

The census of Jordan (Department of Statistics (Jordan), 2015) asks households, "Does your house have sanitation connected to a public network?"

Figure A3: Locations from UN Cities data in our sample



Note: *Data from the* UN DESA *World Urbanization Prospects* (UN DESA *Population Division, 2018), which is a census of all cities that had a population 300,000 or more in 2014. We consider cities in Brazil, Colombia, South Africa, Jordan and Tanzania.*

The Tanzanian census (National Bureau of Statistics (Tanzania), Office of the Chief Government Statistician, 2012) reports a count of households responding yes to, "Does your house have a flush toilet connected to a piped sewer system?"

Table A19 tabulates more detailed responses on households' sewage disposal for tracts used in the estimation sample from Brazil, Jordan, South Africa, and Tanzania. Most households in the sample were connected to a sewer with cesspits and soak pits being the most common alternative to a sewer connection. Colombia did not provide detailed information on the type of sewage disposal at the manzana level. For Colombian tracts used in the estimation sample, 548,115 households reported being connected to a sewer while 180,441 households were not.

Table A19: Households by sewage disposal type

Country	Sewer	Cess/soak pit	Septic tank	Other/none
Brazil	3,488,557	855,152	609,014	199,404
Jordan	12,604	6,059	0	235
South Africa	480,059	89,588	12,131	30,490
Tanzania	7,687	85,395	15,435	21,637
Total	3,988,907	1,036,194	636,580	251,766

C. Population density, other outcomes

Using the censuses described in the previous subsection, we calculate population density and literacy measures for all countries, as well as income measures for Brazil and South Africa. We also collect measures for access to piped water, electricity garbage service, and whether households are in apartment buildings.

For all countries but Jordan, population density is the full count of people divided by tract area. For Jordan, it is the full count of households divided by tract area.

Income is measured monthly in both Brazil and South Africa. Brazil reports the average nominal monthly income of head of household by setore. South Africa reports counts of individuals in 12 income buckets, one of which is "no income." Respondents are asked to consider gross monthly income (pre-tax and including all possible income sources). We approximate the average monthly income for each small area by assigning each income bin the midpoint of the income range for the same bin, then using the reported count to calculate the mean. Colombia, Tanzania, and Jordan do not report income for the granular geographies we use in our analysis.

Literacy is directly reported in the Brazilian and Tanzanian censuses. We do not observe literacy directly in Colombia and South Africa. However, each of these countries releases educational attainment data: Colombia reports a count of persons who have completed *some* amount of primary school (as well as counts for secondary, college/technical, and graduate); South Africa discloses even more granular educational attainment data. We use completion of any primary school, or higher, as a proxy measure for literacy in these countries. We do not observe literacy in Jordan. Because Brazil reports only on literacy, we cannot consider more detailed measures of educational attainment.

Across countries, we use the most comparable census question to measure the share of households with access to piped water. For Brazil, we calculate the share of households reported as being "connected to a water supply from a general network." For Colombia, we calculate the share of households reported as being "with water service." For Jordan, we calculate the share of households reported as having "drinking water from a public network." For South Africa, we infer the share of households with access to piped water by using the number of households reported as having "no access to piped (tap) water." For Tanzania, we calculate the share of households reported as having "piped water into dwelling."

We also calculate the share of households with garbage collection in every country except for Jordan using the census counts. Brazil reports the number of households "with collected garbage." Colombia reports the number of households "with garbage collection service." South Africa reports the number of households with "refuse removed by a local authority/private company." Tanzania reports the number of households with "refuse regularly collected" or "refuse irregularly collected." We use these counts to calculate the share of households with garbage collection service.

We also estimate the share of households with access to electricity in every country except for Jordan. Brazil and Colombia report the number of households "with electricity." South Africa reports the number of households with "energy from electricity." Tanzania reports the number of households where the source of lighting is "electricity (Tanesco/Zeco)."

Finally, Brazil, Colombia, and South Africa report whether households are in an apartment building. Brazil and Colombia report the number of households of the "apartment type." South Africa reports whether a household is a "flat or apartment in a block of flats."

D. Drainage basins

We construct drainage basins from two digital elevation maps (DEMs) using ArcGIS tools for this purpose. The first DEM derives from Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) (NASA/METI/AIST/Japan

Spacesystems and US/Japan ASTER Science Team, 2019) and the second from the Shuttle Radar Topography Mission (SRTM) (NASA JPL, 2013). We rely primarily on the ASTER DEM, but consider SRTM for robustness checks.

These data report the elevation of most of the Earth's surface at a spatial resolution of about 30m^2. Using these digital elevation maps, we draw all drainage basins within a 75KM radius of the center of each city using a utility available for this purpose as part of ArcGIS. We identify the drainage basin containing the center of each city using coordinates in the UN DESA World Urbanization Prospects data. These are the central basins, and our research design is organized around comparisons of neighborhoods on opposite sides of the boundaries of these basins.

Figure 3 illustrates basin boundaries for Cascavel, Brazil. This is an empirical analog of the basins drawn in figure 2. Both DEMS are constructed by looking down at the Earth's surface from satellites, and so both sometimes confuse treetops and roofs with ground level. Because ground level elevation is what is relevant for our exercise, this raises the possibility that we mismeasure basin boundaries. ASTER is based on longer wavelength radiation that is better able to penetrate treetops and roofs, and so is a better measure of ground level elevation in theory. Thus, the ASTER data is our primary basis for constructing drainage basin boundaries, and we rely on the SRTM data primarily for robustness checks. A comparison with LIDAR data shows that average error of ASTER is about 4m in four small study areas, although SRTM is about the same. (Uuemaa et al., 2020).

Appendix C Small variance approximation of the MTE model

Our data reports tract averages, $(\overline{Y}_j, \overline{D}_j, \overline{W}_j)$, and the variance of W_{ij} within each tract j. Using these data, we write parcel level variables as the sum of tract averages and residuals:

$$W_{ij} = \begin{bmatrix} \overline{X}_j + \sigma\epsilon_{ij}^X \\ \overline{Z}_j + \sigma\epsilon_{ij}^Z \end{bmatrix} = \overline{W}_j + \sigma\epsilon_{ij},$$

where σ^2 is a scaling factor for the residuals, and $\epsilon_{ij} = [\epsilon_{ij}^X, \epsilon_{ij}^Z]$ is a vector of residuals such that the within-tract means satisfy $E(\epsilon_{ij}|j) = 0$ and the tract level variance-covariance matrix for ϵ_{ij}, $Var(\epsilon_{ij}|j) = R_j$ exists. Accordingly, we can express $Var(W_{ij}|j)$ as

$$Var(W_{ij}|j) = \Sigma_{W,j} = \sigma^2 \left[\begin{array}{cc} R_j^{XX} & R_j^{XZ} \\ R_j^{XZ} & R_j^{ZZ} \end{array} \right]. \qquad \text{(Appendix C.1)}$$

Assume that the parcel level propensity score $p(W_{ij}) = \Pr(D_{ij} = 1|W_{ij})$ is three times continuously differentiable. Then the second-order Taylor expansion of $p(W_{ij})$ with respect to σ around $\sigma = 0$ is,

$$p(W_{ij}) = p(\overline{W}_j) + \nabla p(\overline{W}_j)\sigma\epsilon_{ij} + \frac{\sigma^2}{2}\epsilon_{ij}' \nabla^2 p(\overline{W}_j)\epsilon_{ij} + O(\sigma^3\|\epsilon_{ij}\|^3). \qquad \text{(Appendix C.2)}$$

In this equation, $\nabla p(W_{ij})$ denotes the $1 \times dim(W_{ij})$ vector of partial derivatives of $p(W_{ij})$ with respect to W_{ij} evaluated at W_{ij}, $\nabla^2 p(W_{ij})$ denotes the corresponding Hessian matrix, and $O(\sigma^3\|\epsilon_{ij}\|^3)$ is the residual of the expansion that vanishes as $\sigma^3\|\epsilon_{ij}\|^3 \to 0$.[15] Taking the expectation conditional on i being in the tract j, we have

$$\begin{aligned} p_j \equiv E[D_{ij}|j] = E[p(W_{ij})|j] &= p(\overline{W}_j) + \frac{\sigma^2}{2} E\left[\epsilon_{ij}' \nabla^2 p(\overline{W}_j)\epsilon_{ij} \middle| j \right] + O_j(\sigma^3) \\ &= p(\overline{W}_j) + \frac{\sigma^2}{2} tr\left(\nabla^2 p(\overline{W}_j) \cdot R_j \right) + O_j(\sigma^3), \end{aligned} \qquad \text{(Appendix C.3)}$$

where we let $O_j(\sigma^3) \equiv E[O(\sigma^3\|\epsilon_{ij}\|^3)|j]$, invoking the assumption of finite third-order moments of ϵ_{ij}. The term $\nabla p(\overline{W}_j)\sigma\epsilon_{ij}$ drops out because $E[\epsilon_{ij}|j] = 0$.

We observe $\overline{D}_j$, the share of households in tract j with sewer access. This measures p_j. We also observe $\overline{W}_j$, and $\Sigma_{W,j} = \sigma^2 R_j$. Therefore, given a functional form for the propensity score $p(\cdot)$ and using the tract level observations, we can use (Appendix C.3) to approximately estimate the parcel level propensity score function $p(\cdot)$, where the approximation error is a higher order term of σ than the variance σ^2.

[15] Equation (Appendix C.2) shows the advantage of the representation of $\Sigma_{W,j}$ in equation (Appendix C.1). By introducing the scaling factor σ^2, we facilitate a univariate Taylor series expansion in (Appendix C.2). The same comment applies to equation (Appendix C.5) below.

We now consider the local IV regression (3). Substituting equation (Appendix C.2) into (3) gives,

$$Y_{ij} = X_{ij}'\beta_0 + \left[p(\overline{W}_j) + \nabla p(\overline{W}_j)\sigma\epsilon_{ij} + \frac{\sigma^2}{2}\epsilon_{ij}'\nabla^2 p(\overline{W}_j)\epsilon_{ij} + O(\sigma^3\|\epsilon_{ij}\|^3)\right](\overline{X}_j + \sigma\epsilon_{ij}^X)'\cdot(\beta_1-\beta_0)$$
$$+\,\phi\left(p(\overline{W}_j) + \nabla p(\overline{W}_j)\sigma\epsilon_{ij} + \frac{\sigma^2}{2}\epsilon_{ij}'\nabla^2 p(\overline{W}_j)\epsilon_{ij} + O(\sigma^3\|\epsilon_{ij}\|^3)\right) + U_{ij}. \qquad \text{(Appendix C.4)}$$

Next, we take a second-order Taylor expansion of $\phi(\cdot)$, take the conditional expectation given i belonging to tract j as in (Appendix C.3), and finally, absorb higher-order terms in $O_j(\sigma^3)$. This gives,

$$\overline{Y}_j = \overline{X}_j'\beta_0 + p(\overline{W}_j)\overline{X}_j(\beta_1-\beta_0) + \sigma^2\left[\nabla p(\overline{W}_j)R_j^{WX} + \frac{1}{2}tr\left(\nabla^2 p(\overline{W}_j)\cdot R_j\right)\overline{X}_j'\right](\beta_1-\beta_0)$$
$$+\,\phi(p(\overline{W}_j)) + \phi'(p(\overline{W}_j))\frac{\sigma^2}{2}tr\left(\nabla^2 p(\overline{W}_j)\cdot R_j\right)$$
$$+\,\phi''(p(\overline{W}_j))\frac{\sigma^2}{2}\nabla p(\overline{W}_j)R_j(\nabla p(\overline{W}_j))' + O_j(\sigma^3) + \eta_j, \qquad \text{(Appendix C.5)}$$

where $R_j^{WX} = E[\epsilon_{ij}\epsilon_{ij}^{X\prime}|j] = \begin{bmatrix} R_j^{XX} \\ R_j^{XZ} \end{bmatrix}$ is a submatrix of R_j.

Finally, assume,

$$\phi(p) = \alpha_0 + \alpha_1 p + \frac{1}{2}\alpha_2 p^2. \qquad \text{(Appendix C.6)}$$

Using parameter estimates from equation (Appendix C.3) we can evaluate $p(\overline{W}_j)$ and its derivatives at $\overline{W}_j$. This means that we can define three observable scalars,

$$k_{1j} \equiv \left[\nabla p(\overline{W}_j)R_j^{WX} + \frac{1}{2}tr\left(\nabla^2 p(\overline{W}_j)\cdot R_j\right)\overline{X}_j'\right]$$
$$k_{2j} \equiv tr\left(\nabla^2 p(\overline{W}_j)\cdot R_j\right)$$
$$k_{3j} \equiv \nabla p(\overline{W}_j)R_j(\nabla p(\overline{W}_j))'.$$

Substituting into the structural equation (Appendix C.5), we have

$$\overline{Y}_j = \overline{X}_j'\beta_0 + p(\overline{W}_j)\overline{X}_j(\beta_1-\beta_0) + \sigma^2(\beta_1-\beta_0)k_{1j} \qquad \text{(Appendix C.7)}$$
$$+\,\alpha_0 + \alpha_1\left[p(\overline{W}_j) + \frac{\sigma^2}{2}k_{2j}\right]$$
$$+\,\alpha_2\left[\frac{1}{2}p(\overline{W}_j)^2 + p(\overline{W}_j)k_{2j} + \frac{\sigma^2}{2}k_{3j}\right] + O_j(\sigma^3) + \eta_j.$$

We arrive at equation (8) and (9) in the main text by assuming the linear probability model, equation (7), in both (Appendix C.3) and (Appendix C.7), noting $k_{2j} = 0$, and including approximation error in the regression residual.

Our estimand of interest is the SATE. In this case it is given by,

$$\text{SATE} = \overline{X}'(\beta_1 - \beta_0) + \alpha_1 + \frac{1}{2}\alpha_2, \tag{Appendix C.8}$$

where $\overline{X}$ is the sample average of X_j. We estimate SATE by plugging in the OLS estimator of the relevant coefficients obtained from (Appendix C.7).

To obtain a standard error estimate for our SATE estimator, we express our estimation of γ and (α,β) using method of moments. Denote the regressor vector (column vector) of equation (Appendix C.7) by,

$$S_j(\gamma) = \left(X_j', \quad (W_j'\gamma) \cdot X_j' + \sigma^2 \gamma' R_j^{WX}, \quad W_j'\gamma, \quad \tfrac{1}{2}(W_j'\gamma)^2 + \tfrac{\sigma^2}{2}\gamma' R_j \gamma\right)'$$

and the coefficient vector by $\theta = (\beta_0', \beta_1' - \beta_0', \alpha_1, \alpha_2)'$ with α_0 absorbed into the intercept parameter. Then (θ,γ) is the solution to

$$E\begin{bmatrix} S_j(\gamma)(\overline{Y}_j - S_j(\gamma)'\theta) \\ m_j(\gamma) \end{bmatrix} = 0, \tag{Appendix C.9}$$

where $m_j(\gamma)$ is the first-order condition for γ in the OLS estimation of (7). Substituting sample analogs gives our estimators,

$$\frac{1}{n}\sum_{j=1}^{n}\begin{bmatrix} S_j(\widehat{\gamma})(\overline{Y}_j - S_j(\widehat{\gamma})'\hat{\theta}) \\ m_j(\widehat{\gamma}) \end{bmatrix} = 0. \tag{Appendix C.10}$$

If we now expand the sample moment conditions around the true parameter values, (θ, γ), we have

$$0 = \frac{1}{n}\sum_{j=1}^{n}\begin{bmatrix} S_j(\gamma)(\overline{Y}_j - S_j(\gamma)'\theta) \\ m_j(\gamma) \end{bmatrix} + \begin{pmatrix} \widehat{\nabla}_{1,\theta} & \widehat{\nabla}_{1,\gamma} \\ O & \widehat{\nabla}_{2,\gamma} \end{pmatrix} \cdot \begin{pmatrix} \hat{\theta} - \theta \\ \widehat{\gamma} - \gamma \end{pmatrix} + \text{remainder}, \tag{Appendix C.11}$$

where $\widehat{\nabla}$'s are the derivative matrices of the sample first order conditions.

Multiplying both sides by $\sqrt{n}$, solving for parameters, and letting $n \to \infty$, we obtain the following asymptotic approximation;

$$\begin{aligned}\sqrt{n}\begin{pmatrix} \hat{\theta} - \theta \\ \widehat{\gamma} - \gamma \end{pmatrix} &= -\begin{pmatrix} \nabla_{1,\theta} & \nabla_{1,\gamma} \\ O & \nabla_{2,\gamma} \end{pmatrix}^{-1} \cdot \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\begin{pmatrix} S_j(\gamma)(\overline{Y}_j - S_j(\gamma)'\theta) \\ m_j(\gamma) \end{pmatrix} \\ &\to_d \mathcal{N}(0, \nabla^{-1}\Sigma(\nabla^{-1})'),\end{aligned} \tag{Appendix C.12}$$

where

$$\nabla = E\begin{pmatrix} -S_j(\gamma)S_j(\gamma)' & \nabla_\gamma S_j(\gamma)\eta_j - S_j(\gamma)\theta'\nabla_\gamma S_j(\gamma) \\ O & \nabla_\gamma m_j(\gamma) \end{pmatrix} \qquad \text{(Appendix C.13)}$$

$$= E\begin{pmatrix} -S_j(\gamma)S_j(\gamma)' & -S_j(\gamma)\theta'\nabla_\gamma S_j(\gamma) \\ O & \nabla_\gamma m_j(\gamma) \end{pmatrix}. \qquad \text{(Appendix C.14)}$$

The second equality holds because η_j is a regression residual with mean zero conditional on the regressors in (Appendix C.7). Σ is the variance covariance matrix of the moment conditions.

$$\Sigma = E\left[\begin{pmatrix} S_j(\gamma)(\overline{Y}_j - S_j(\gamma)'\theta) \\ m_j(\gamma) \end{pmatrix} \cdot \begin{pmatrix} S_j(\gamma)(\overline{Y}_j - S_j(\gamma)'\theta) \\ m_j(\gamma) \end{pmatrix}'\right]$$

If we specify the linear probability model for (7), then we have $m_j(\gamma) \equiv W_j\nu_j = W_j(\overline{D}_j - W_j'\gamma)$, and $\nabla_\gamma m_j(\gamma) = -W_jW_j'$. Under the same assumption, the derivative matrix of $S_j(\gamma)$ is

$$\nabla_\gamma S_j(\gamma) = \begin{pmatrix} O_{\dim(X)\times\dim(\gamma)} \\ X_jW_j' + \sigma^2(R_j^{WX})' \\ W_j' \\ (W_j'\gamma)W_j' + \sigma^2\gamma' R_j \end{pmatrix}.$$

We estimate the asymptotic variance of (Appendix C.12) by plugging in $\widehat{\gamma}$ in place of γ and replacing the expectation by the sample average for both ∇ and Σ terms,

$$\widehat{\nabla} = \frac{1}{n}\sum_{j=1}^{n}\begin{pmatrix} -S_j(\widehat{\gamma})S_j(\widehat{\gamma})' & -S_j(\widehat{\gamma})\hat{\theta}'\nabla_\gamma S_j(\widehat{\gamma}) \\ O & \nabla_\gamma m_j(\widehat{\gamma}) \end{pmatrix},$$

$$\widehat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}\begin{pmatrix} S_j(\widehat{\gamma})(\overline{Y}_j - S_j(\widehat{\gamma})'\hat{\theta}) \\ m_j(\widehat{\gamma}) \end{pmatrix} \cdot \begin{pmatrix} S_j(\widehat{\gamma})(\overline{Y}_j - S_j(\widehat{\gamma})'\hat{\theta}) \\ m_j(\widehat{\gamma}) \end{pmatrix}'$$

Focusing on the first block element of $\widehat{\nabla}^{-1}\widehat{\Sigma}(\nabla^{-1})'$ gives the asymptotic variance estimate for $\sqrt{n}(\hat{\theta} - \theta)$.

Since the SATE estimator can be expressed as $\widehat{\text{ATE}} = a'\hat{\theta}$ with $a = (\mathbf{0}', \overline{X}', 1, 1)'$, its asymptotic variance can be obtained by the asymptotic variance of $\hat{\theta}$ sandwiched by a' and a.