# Sewers and Urbanization in the Developing World*

preliminary

Sean McCulloch,† Matthew Schaelling ‡ Matthew A. Turner §

19 August 2024

Abstract: *We investigate the effects of sewer access on neighborhood characteristics in developing world cities. Because it is more difficult to move sewage uphill than downhill, otherwise similar neighborhoods on opposite sides of drainage basin divides may face different costs of sewer access. We exploit this intuition to identify the effect of sewer access by comparing outcomes for neighborhoods on opposite sides of drainage basin divides. We estimate the effect of sewer access on census tract level population density, education, and income for Brazil, Colombia, South Africa, Jordan and Tanzania. On average, sewer access has a large effect on population density and almost none on demographics. These estimates imply that in many cities, completing central city sewer networks increases center city access by as much as large increases in transportation infrastrucutre.*

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 for seminar participants at the University of Wisconsin. 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.

# 1 Introduction

We investigate the effects of sewer access on neighborhood characteristics in developing world cities. Our estimates are based 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 neighborhoods on opposite sides of drainage basin divides face different costs of sewer access. We use this intuition, and census tract level data to estimate the effects on population density, education, and income of treating census tracts with more sewer access. We identify treatment effects by comparing rates of sewer access and outcomes for neighborhoods on opposite sides of drainage basin divides in Brazil, Colombia, South Africa, Jordan and Tanzania. Our estimates provide a basis for evaluating the impact of changes to sewer access on developing world cities.

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' "Millenium Development Goals". Yet, many cities also lack decent roads, sufficient public transit, adequate schooling, and reliable electricity. Trade-offs between these services must be evaluated and made. By providing estimates of the effects of sewer access on slums, we hope to inform policy makers facing such trade-offs.

Sewers are probably also important for their indirect effects. 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 will bring new evidence to bear on this question. Does sewer access lead to denser neighborhoods that can accommodate more of the rural poor? Alternatively, does sewer access lead to nicer, but less dense neighborhoods that house affluent newcomers by displacing poor incumbents? Answering these questions is of immediate relevance to development policy. Moreover, if spatial equilibrium reflects the balancing of agglomeration with congestion forces, understanding the importance of sewers for mitigating the costs of urban density is fundamental to understanding the equilibrium.

The effects 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.

## 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.

Studies of sewers are rarer. Alsan and Goldin (2019) study late 19th century Boston and find a large reductions in infant mortality from a collection of policies to keep sewage away from drinking water. These policies did not affect all places equally, and Alsan and Goldin (2019) also find no evidence that people sorted into places with better water supplies 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. (2022a) explicitly considers the relationship between sewer construction and urban development. Coury et al. (2022a) 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.

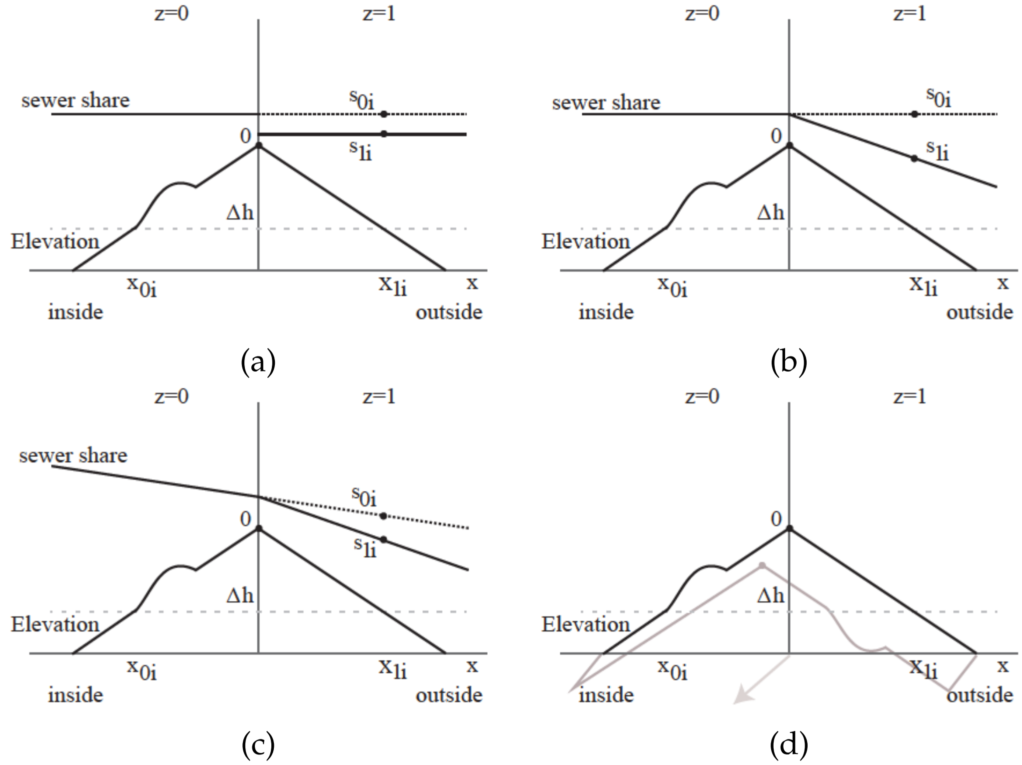
Summing up, the available evidence suggests that sewer access has beneficial effects on cities and neighborhoods. The evidence for more specific effects is thin and based on 19th century US cities. Coury et al. (2022b) find that sewer access leads to large increases in land prices in late 19th century Chicago. It is natural to suspect that these prices increases were associated with an increase in density. If we extrapolate from water supplies to residential sewer service, the finding in Alsan and Goldin (2019) that people do not sort on the basis of local variation on municipal water quality suggest that we should not expect people to sort on the basis of sewer access. 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

### A *Drainage basin divides and the cost of sewers*

The movement of wastewater in sewers is special in two regards. It 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 1:400 solids

Figure 1: Identifying treatment effects around a stylized basin boundary

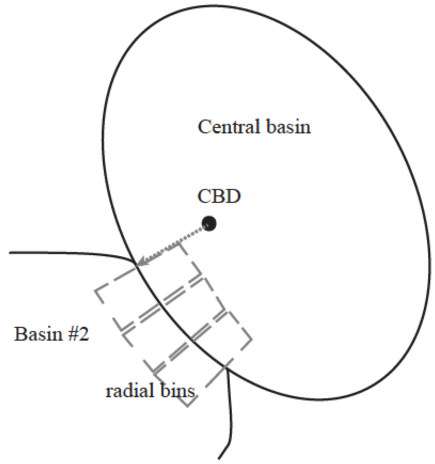



Note: *Elevation and sewer percentage profile in the neighborhood of a drainage basin divide. The basin divide is at the top of the hill, at* $x = 0$*. Displacement left is 'inside' and towards the nearest established sewer system. Displacement right is 'outside' and wastewater in this region must travel 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) Same as (b) but* $x$ *displacement has an independent effect on sewer access. (d) Illustration of variation in elevation independent of* $x$*.*

settle out 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). Also, unlike commuters, wastewater only travels away from a residence. Thus, commuting should respond symmetrically to elevation change on the outbound and inbound trips, but sewers should respond asymmetrically. For household wastewater, only elevation gain outbound is costly.

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 perpendicular to a drainage basin divide at $x = 0$. The region to the left of $x = 0$

Figure 2: Illustration of basins, segments, radial bins, and 'inside' indicator




Note: *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 segment 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'. To construct these bins, we divide the central basin divide into two kilometer long intervals, starting from the pint on the basin divide nearest the city center. A $k$ km radial bin is defined as the area within $k$ kilometers of one such interval.*

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 travelling 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, the grade of the sewer pipe can diverge from that of the ground above. Recalling that a sewer needs a grade of only 1:200, this means that burying a sewer to a depth of eight feet instead of two can allow an extra 1200 feet of horizontal travel. Second, if the topography allows, following an indirect route approximately following an elevation contour to reach the inside of the basin allows the substitution of downhill, horizontal travel for climbing. Third is the construction of pumping facilities, which in turn requires the availability of electric or fossil fuel powered pumps.[1]

[1]If the land outside the central basin is sufficiently valuable, there is also the possibility of building a new sewer network to service the relevant drainage basin.

Summing up, for places on the outside of a drainage basin divide, the cost of reaching the central city sewer network should increase rapidly with the horizontal and vertical distance that sewage must cover to reach the basin divide (from which it can drain downhill to the central city sewer network). Conversely, for places on the inside of the basin divide, horizontal and vertical displacement 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 a sample of locations close to a drainage basin divide, location inside or outside the basin is a source of quasi-random variation in the cost of sewers. Our research design is organized around using this cost shock to estimate the treatment effects for sewer access.

The four panels of figure 1 inform the exercise of translating this intuition into an econometric specification. The horizontal axis, $x$, is displacement along an axis perpendicular to a drainage basin divide. 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 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 1 for $x$ outside, and 0 otherwise. Let $s$ indicate the percentage of households in a location with sewer access, and define $\Delta(x)$ 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 $\mathbb{1}(\text{Outside})(x)\Delta(x)$ are the horizontal and vertical displacement required to reach the inside of the central drainage basin from location $x$. We will consider $\mathbb{1}(\text{Outside})$, $\mathbb{1}(\text{Outside})(x)x$ , and $\mathbb{1}(\text{Outside})(x)\Delta(x)$ as instruments.

Our measure of sewer access is the percentage 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 usually at least one half kilometer square. 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 percentage, $s$ does not depend on $x$, except at the basin divide, where the cost of sewer access increases, and the percentage of houses reporting access to a sewer declines as a step function.

We would like to know how an outcome $Y$ depends on sewer access. Panel (a) suggests that we do this by estimating the step down in sewer access at $x = 0$ and comparing it to the change in $Y$ that occurs around this step.

In figure 1(a), census units outside are all treated, and census units inside are all untreated controls. It is hard to have a strong prior about the spatial scale over which the basin divide cost shock will operate. There are a three reasons to think that it will not operate as sharply as illustrated in panel (a). First, the area near a drainage basin divide is often quite flat, a few tens or hundreds of feet on the outside of the basin divide may involve only a foot or two of drop. This can be accommodated easily by burying the sewer a little deeper. Second, in practice, we will measure the locations of basin divides imprecisely, so measurement error should smooth out the empirical counterpart of figure 1(a). Third, even if the area near the drainage basin divide is not flat, the land may be valuable enough that residents and developers extend the sewer network outside the central basin.

In any of these cases, we expect sewer access to decline continuously as locations are progressively further outside the the central basin, not discretely. This case is illustrated in figure 1 (b). The intuition behind panels (a) and (b) is similar, but the implied econometric model is not. Panel (a) can be described by a discrete instrument, $(z = 0,1)$, and a discrete treatment,$(s = s_0, s_1)$. In panel (b) the cost shock increases in distance, as does the resulting change in sewer percentage, so instrument and treatment are both continuous.

Figure 1(c) illustrates one of the main challenges to our identification strategy. By construction, the drainage basin encircles the center of the city it contains. Thus, displacement inside is towards the city center and conversely. As we move close to the center, we expect land to become more valuable and more intensively developed. Assuming the cost shock to sewers operates continuously, as illustrated in figure 1(c) we expect a steady decline in sewer percentage as we move from left to right, away from the city center, with a trend break and more rapid decline once we cross the basin divide.

We now turn to the role of of vertical displacement. In all panels of figure 1, there is variation in elevation independent of $x$. If we are concerned that elevation has an independent effect on sewer access and outcomes, this can be identified independent of the effect of horizontal displacement. Moreover, the logic we began with suggests that the effect of elevation on the cost of sewers should be different outside the basin than inside. In the drawing of figure 1, vertical and horizontal displacement are nearly identical, and so the same logic that suggests using horizontal displacement inside and outside to identify the effect of sewers on outcomes also applies to vertical displacement.

Figure 1(d) illustrates a final point about our identification strategy. Unlike the first three panels, our data will lie on 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 displacement to have a larger effect on sewer access outside the basin divide than inside (where we expect the effect of elevation to be small or zero).

It is at least possible that the basin boundary is an important geological feature and that crossing it affects sewer percentage and outcomes because it impedes the movement of people and goods along with wastewater. We show below that this is probably not the case. On average, basin divides are small enough geological features to often be unnoticeable. Second, there are many places where sewer percentage does not vary across the basin divide. These places fall into two categories. Those that are sufficiently undesirable that they are unsewered on both sides of the basin divide, and those that are sufficiently desirable that they are completely sewered on both sides of the divide. For both classes of location, we compare outcomes across the basin divide and find no change. This suggests that crossing the basin divide does not have an effect on outcomes independent of its effect on sewer access.

Finally, we turn to the problem of defining an empirical analog to the illustrations in figure 1, and in particular to defining the notion of 'inside' and 'outside'. We define a census tract as 'inside' or 'outside' depending on whether its area weighted centroid is on the same side of the closest basin divide as the central business district contained in all of our central basins.[2] Because cities must have a population of at least 300k to be included in our sample, all central

[2]We experimented with using lights-at-night weighted centroid and found that it not have an important effect on our results.

cities have at least some sewer service. Thus, this definition guarantees that an inside census tract can drain to a central city sewer network.

We exclude from our sample any census tract for which the closest basin divide is not the central basin divide for one of the cities the UN DESA data. Our sample includes only census tracts for which the closest basin divide describes the boundary of one of our cities.

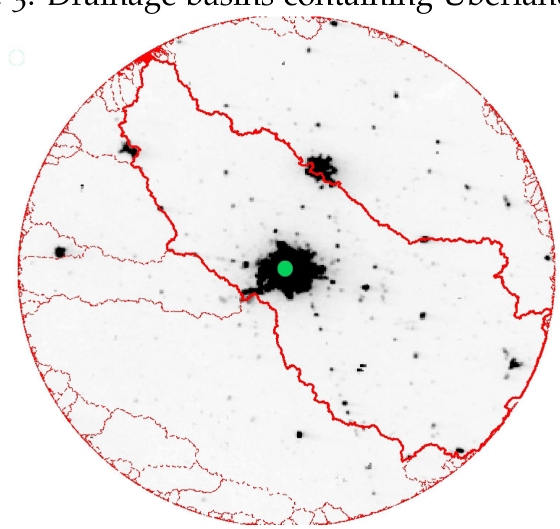
This raises the question of how should we define inside and outside when two central drainage basins are adjacent? This is occurs in a about XX% of our sample tracts. There are two natural solutions. The first is to exclude such tracts from our sample. Alternatively, for tracts for which the closest basin divide segment divides two central basins, define 'inside' on the basis of the closest of the two city centers. We experiment with both strategies and our qualitative results are robust to either definition. However, when this situation arises, the more remote CBD is three times as far away as the closer one on average. Therefore, while this notion of 'inside' and 'outside' can be ambiguous in theory, it is rarely ambiguous in practice.

A problem may 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 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.

The drainage basins that contain our cities are sometimes far too small to contain a meaningful share of the city's population (recall the UN DESA data reports on cities with a population above 300,000). This is a particular problem for cities near the coast, where the algorithm that draws drainage basins tends to construct small basins. To see why this creates a problem consider two 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 10 kilometers. Pooling these two types of basin divides complicates the interpretation of displacement perpendicular to the basin divide. To resolve these problems, we define the central basin as the union of drainage basins that intersect a disk of 1km radius centered on the CBD. This guarantees that no point on the central basin divide is closer to the CBD than 1km.

# 4 Data

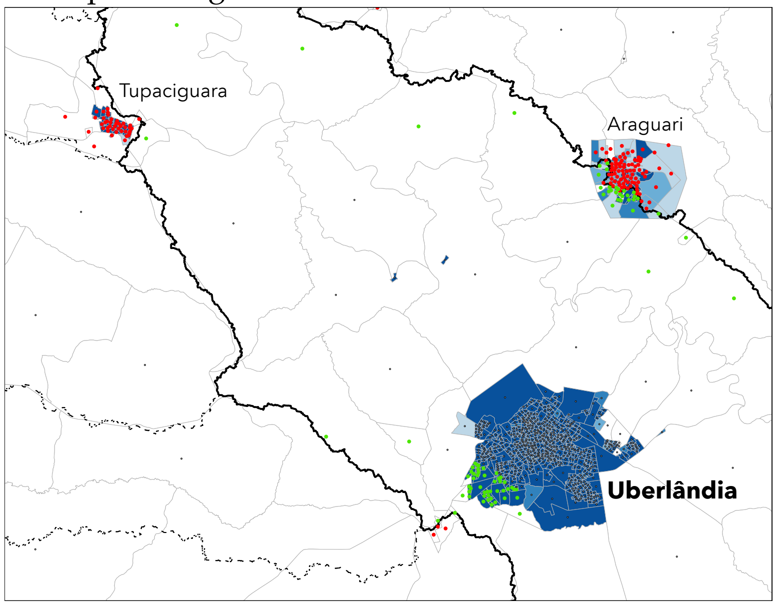
Figure 3: Drainage basins containing Uberlandia, Brazil



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 Uberlandia, Brazil. The green dot indicates the center of the city. VIRRS lights at night shows city extent. The disk has a radius of 75km.*

We use the UN DESA World Urbanization Prospects data to identify the centers of all cities that have a population of 300,000 or above in 2018 (DESA Population Division, 2018). We restrict attention to the cities in countries where we have census data and maps: Brazil, South Africa, Tanzania, Jordan and

Figure 4: Sewer percentage and 'inside/outside' around Uberlandia, Brazil




Note: *Close up of Uberlandia, Brazil. Darker blue indicates a larger percentage 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 (black), as are centroids that are more than 4km 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, centroids more than 69km from the city center are also excluded from the sample.*

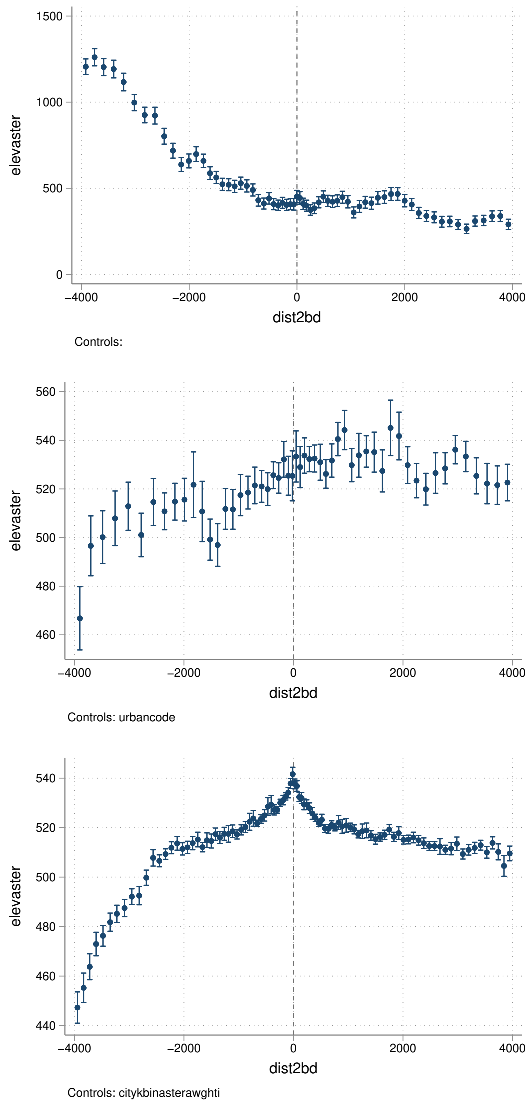
Colombia.

We download the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) (NASA/METI/AIST/Japan Spacesystems and U.S./Japan ASTER Science Team, 2019) digital elevation map. These data report the elevation of most of the Earth's surface at a spatial resolution of about 30m$^2$.

We next clip of a circle of radius 75km surrounding each of the cities in our sample. This done, we draw all drainage basins within a 75km radius of 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 our research design is organized around comparisons of neighborhoods on opposite sides of the divides that define these basins.

Figure 3 illustrates basin boundaries around Uberlandia, Brazil, and is an empirical analog of figure 2. Red dashed lines indicate the boundaries of all

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; raw data (top), net of segment mean (middle), and net of radial bin mean (bottom). Note that the y-axis scale is much larger on the top panel and about the same for the bottom two.*

drainage basins, and solid red shows the boundary of the central basin. Shading is based on lights at night and shows the scale of the city relative to the various basins. Notice that our algorithm smetimes constructs incoherent basins at the edge of the map disc. For this reason, we exclude the region within six kilometers of the edge of these discs, or conversely, more than 69 km from the center of the city.

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 Colombia, Tanzania, Jordan, and South Africa. Appendix B provides more detail about our census data.

Figure 4 is a heat map illustrating the incidence of sewer access for the Brazilian city of Uberlandia. Polygons describe the extent of census tracts, with darker blue indicating a larger percentage of households reporting sewer access. Basin boundaries are black lines. Dots indicate census centroids. Census centroids are red if they are inside the central basin and green outside. Grey indicates centroids excluded from our sample.

Our data so far describe census tract level sewer percentage and other outcomes, and the locations of tract centroids relative to drainage basin divides. It remains to calculate the vertical distance between each tract and the central basin divide. For this purpose, 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, and within 2km of the divide. We then calculate the vertical rise required to reach the basin divide, $\Delta(x)$ in our earlier notation, as the elevation difference between the target tract centroid and this radial bin maximum. That relative elevation is determined at the $\pi$-bin level motivates our attention to errors clustered at this level.

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 4km of the central basin divide, (2) closer to the central basin divide than to any other central basin divide, (3) within 69 km of the city center. We assign sewer percentage, population density and other outcomes on the basis the relevant census survey data, and vertical distance to basin divide on the basis of the elevation of the highest tract centroid in the relevant radial bin. Table 1 describes out sample. Appendix A provides

Table 1: Descriptive statistics

| | CBD Basin + 4km | $\pm$ 4km Basin Divide |
|---|---|---|
| Num cities | 86 | 85 |
| Mean area cbd basin (kmsq) | 6,613 | 1,494 |
| Num segments | 1,103 | 582 |
| Num radial-bins | 9,245 | 3,288 |
| Num tracts | 396,967 | 76,995 |
| Share inside | 0.88 | 0.48 |
| Mean tract area (kmsq) | 1.08 | 0.90 |
| Mean dist to CBD (km) | 22.45 | 13.45 |
| Mean log dist to CBD (m) | 9.44 | 8.83 |
| Mean dist to basin divide (km) | 12.21 | 1.67 |
| Sewer share | 0.67 | 0.69 |
| Pop density (persons per kmsq) | 26,019 | 21,798 |
| Income (pay per month) | 1,582 | 1,513 |

Note:*About 25 tracts per bin. Tracts are about 1km square. Central basins average about 80km square. Population density is high. Quarter acre lots with 4 people per household is about 4000 per km*$^2$.

summary statistics by country.

Figure 5 shows empirical analogs of the elevation profile in figure 1 based on the universe of census tracts within 4km of a central basin divide. Each of the three panels 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, the x-axis. As in figure 1, the basin divide is at $x = 0$, displacement to the left of zero is displacement inside the central basin, and displacement to the right is outside the central basin. The top panel shows unconditional bin means. In this figure, there is no discernible evidence of the expected high point at the basin divide. In the middle panel, we repeat the exercise using tract centroid elevations net of city mean elevation. In this panel, we begin to see the expected high point at the basin divide. In the bottom panel, we repeat the exercise again, but based on tract centroid elevation net of radial-bin mean elevation. We see the expected peak at the basin divide.

Figure 5 illustrates three important features of our data. First, the drainage basin divides are not dramatic geographic features. Comparing the bottom panel of figure 5 to the panels above, we see that the variation in elevation associated with a few kilometers of travel perpendicular to the basin divide is small relative to the variation across cities, or the variation within a city as we travel circumferentially along the basin divide. The bottom panel of figure 5 also shows that the basin divides are small features in an absolute sense. On average, travelling 2km perpendicular to the basin divide, whether inside or outside, involves a descent of about 30m. Thus, the average grade along a 4km path extending from 2km inside the basin divide to 2km outside is about 1:70, close to the threshold at which athletes begin to notice a playing field is sloped.

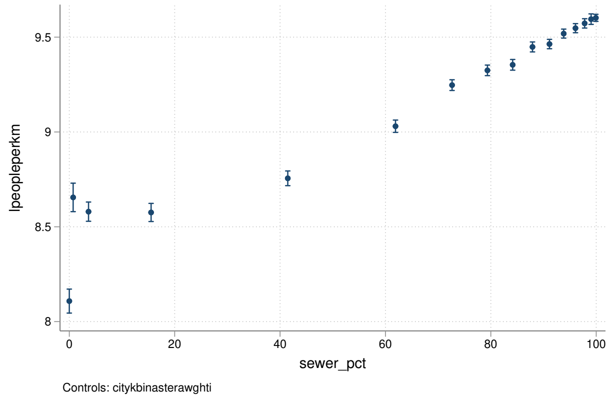
Second, the peak at $x = 0$ visible in the bottom panel of figure 5 is by construction. Basin divides are constructed to lie at local high points. That we cannot see the basin divide in the top two panels indicates that when we are comparing tracts across the basin divides, we are not making the comparisons we intend. It is only once we control for radial bin means that we seem to be comparing tracts that are 'close enough' together that the expected pattern in the data emerges. This motivates our reliance on radial bin level variation in our estimations.

Third, whiskers in the bottom panel describe standard deviation of elevation around the mean, conditional on displacement from the basin divide. One of our instruments is constructed around exactly this variation, and the bottom panel gives a sense for its magnitude.

Figure 6 illustrates variation in our two main variables of interest, the percentage of households in a tract with sewer access and the logarithm of population density. The figure is a binscatter plot, and so the slop reflects means in the raw data. Throughout most of the range of sewer access, the relationship is approximately linear, and the slope indicates an elasticity around three. That is, each 1% increase in sewer access is associated with a 3% increase in population density. Because we expect (and hope) that the assignment of sewer access to census tracts is not done at random, we cannot interpret this slope as a causal effect of sewer access on density. Estimating this casual relationship is the central econometric problem that we address.

Figure 7 provides evidence for the validity of our research design. The top panel is a binscatter plot showing mean distance to the CBD net of radial bin means. As expected, this plot is continuous. This is intuitive, but not guaranteed.

Figure 6: Logarithm of tract Population density vs sewer %




Note: *Mean log population by tract sewer percentage. All tracts within 4km of a basin divide, conditional on radial bin. 100% increase ≈ factor of three for pop. density, i.e. 3% elasticity.*

If we had seen a step or kink in this distance gradient when we crossed the basin divide, it would have indicated a problem with our sampling rule. The bottom panels plot our main outcome, log tract population density, as a function perpendicular displacement the basin divide. In panel (b) we restrict attention to the subset of radial bins where tract mean sewer percentage is above 90% within 2km inside and outside of the basin divide. Panel (c) is similar, but restricts attention to radial bins where tract mean sewer percentage is below 10% within 2km inside and outside of the basin divide. That is, panels (b) and (c) plot how population density changes when we cross a basin divide where sewer percentage does not vary. That population density stays constant across these divides suggests that basin divides do not have an effect on population density independent of sewer percentage.

Our econometric model is organized around two main types of quasi-experimental variation illustrated in figure 8. The first is changes in the level or trend of tract level sewer percentage when we cross the central basin divide. Panel (a) illustrates this variation in our sample. This plot is a binscatter showing the bin of tract percentage sewer access, conditional on radial bin means. The $x$-axis is perpendicular displacement from the basin divide, and the construction of this figure is like that for elevation in the bottom panel of figure

5. This is a 'first-stage' regression. We see a trend break in sewer percentage at $x = 0$, and possibly a small step down. This is consistent with our intuition about the costs of sewer construction. Crossing the basin divide increases the cost of sewers and decreases their prevalence, but the exact functional form of this relationship is not obvious, and may be confounded by an independent effect of elevation or horizontal displacement. In our econometric model, we experiment with using an indicator for outside, $\mathbb{1}(\text{Outside})$, and an indicator for outside interacted with displacement, $\mathbb{1}(\text{Outside})x$, as instruments. The validity of these instruments probably depends on controlling for elevation and horizontal displacement, and we experiment with different specifications using these controls.

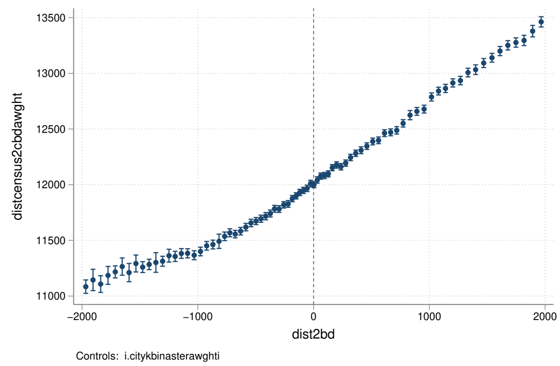
Figure 8(c) is like (a) but reports bin means of tract log population density on the $y$-axis. This is a reduced form regression. This figure is less clear than the corresponding figure for sewer percentage, but population density appears to decline outside the central basin.

Comparing the two figures suggests a magnitude of the treatment effect. The step down in panel (a) is about a 10% decline in sewer percentage. The step down in panel (c) is from about 9.2 to 9.1, this is about a 10% decrease in population density. Dividing, we conclude that a 1% decline in sewer percentage gives about a 1% decline in population density. Our econometric specification allows for richer controls than we use here, and for crossing the basin divide to affect the trends of sewer access and population density. This generally leads to somewhat larger estimates of effect size.

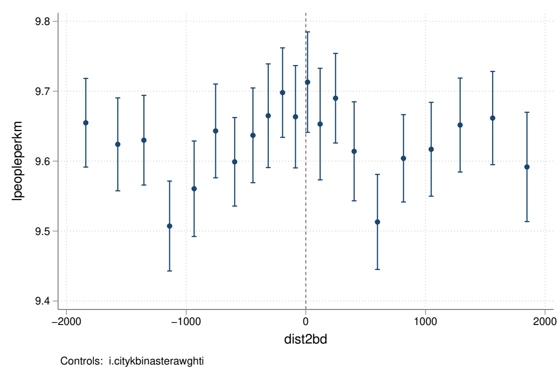
Panel (b) describes our second source of identifying variation. Like panel (a), this figure is a binscatter plot and reports bin mean of tract sewer percentage. Unlike panel (a) bin means are conditional on radial bin means and on linear trend in displacement $x$. Also unlike panel (a), the $x$-axis reports vertical displacement required to reach the basin divide. The region to the left of $x = 0$ gives vertical displacement for tracts on the inside of the basin, and conversely for the region to the right of $x = 0$. Thus, the bin at $x = 10$ reports the mean conditional sewer percentage for tracts 10 feet below the basin divide. Thus, this figure is also in the spirit of a first stage regression, and shows a clear trend break in sewer percentage at $x = 0$.

Panel (d) is the corresponding reduced form. This binscatter plot is constructed in the same way as panel (b) but reports bin mean log population density. This figure also shows a trend break and possibly a step at $x = 0$. Therefore, like panels (b) and (d) suggest that vertical displacement can also be used to identify the effects of sewer percentage on population density.

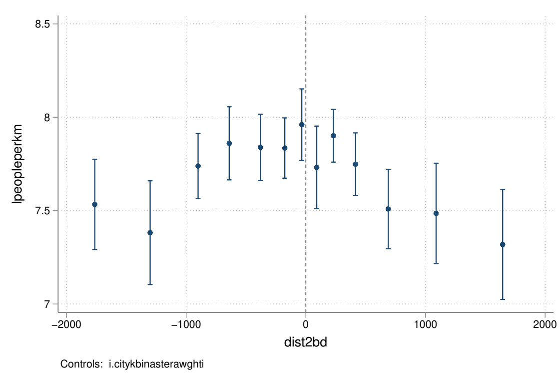
Figure 7: Placebo and balance tests




(a)



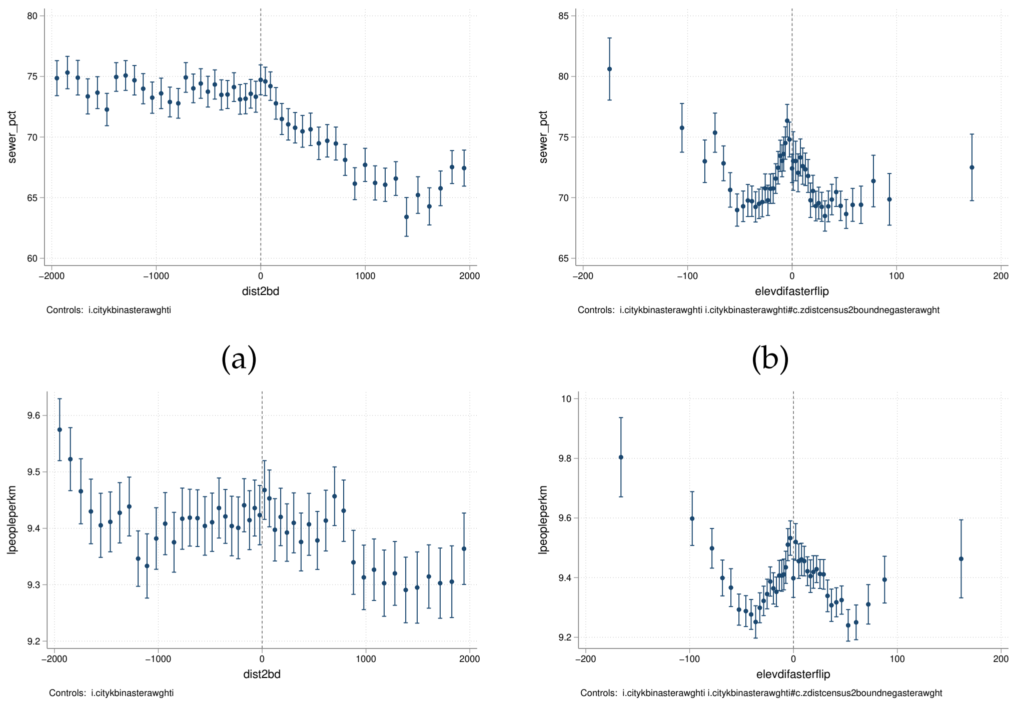

(b)




(c)

Note: *Each panel is a binscatter plot with perpendicular displacement from the central basin divide on the $x$-axis. (a) Bin mean of distance to cbd net of radial bin mean. That this plot is continuous is reassures us that we have not introduced an unintended sampling restriction. (b) Bin mean population density for all radial bins where tract mean sewer percentage is above 90% within 2km of the basin divide. (c) Same as (b) but for radial bins where tract mean sewer percentage is below 10% within 2km of the basin divide. That population density is about constant across portions of the central basin divides where sewer percentage does not vary suggests that the basin divide does not affect population density independent of sewer percentage.*

Figure 8: Identifying variation around a stylized basin boundary




(a) (b) (c) (d)

Note: *All panels are binscatter plots. (a) Mean tract sewer percentage net of radial bin fixed effects as a function of perpendicular displacement. (b) Mean tract sewer percentage as a function of vertical distance to basin divide, net of radial bin mean and a bin specific trend in perpendicular displacement. (c) Mean tract population density net of radial bin fixed effects as a function of perpendicular displacement. (d) Mean tract population density as a function of vertical distance to basin divide, net of radial bin mean and a bin specific trend in perpendicular displacement. (a) and (b) are the analogs of the plot of $s$ in figure 1.*

# 5 Econometric model and results

We would like to estimate the effects of sewer access more precisely than is possible by inspection of figure 8. To proceed, let $i$ index census tracts and $k$ index radial bins. $s_{ik}$ is the percentage, from 0-100, of households in tract $i$ and radial bin reporting sewer access. $y_{ik}$ is the corresponding outcome of interest, most often the logarithm of population density, but sometimes a measure of tract mean education or income. $x_{ik}$ is meters from the tract centroid to basin divide, with displacements inside the basin negative and displacements outside positive. $\Delta E_{ik} \geq 0$ is the vertical rise required to reach the basin divide from the centroid of tract $i$. Note that $\Delta E_{ik}$ is strictly greater than zero for every tract in a radial bin $k$, except for the tract that defines the elevation of the basin divide for that bin. Finally, let $\mathbb{1}(\text{Outside})_{ik}$ be an indicator variable that is one for tracts with centroids outside the central basin, and zero otherwise.

Consider the following three estimating equations,

$$
\begin{aligned}
s_{ik} =& \mathbb{1}(k)_{ik} + \mathbb{1}(k)x_{ik} + A^s \Delta E_{ik} + \\
& \alpha_0 \mathbb{1}(\text{Outside})_{ik} + \alpha_1 \mathbb{1}(\text{Outside})_{ik} x_{ik} + \alpha_2 \mathbb{1}(\text{Outside})_{ik} \Delta E_{ik} + \eta^s_{ik} \quad (1) \\
y_{ik} =& \mathbb{1}(k)_{ik} + \mathbb{1}(k)x_{ik} + A^y \Delta E_{ik} + \\
& \gamma_0 \mathbb{1}(\text{Outside})_{ik} + \gamma_1 \mathbb{1}(\text{Outside}) x_{ik} + \gamma_2 \mathbb{1}(\text{Outside}) \Delta E_{ik} + \eta^y_{ik} \quad (2) \\
y_{ik} =& \mathbb{1}(k)_{ik} + \mathbb{1}(k)x_{ik} + A \Delta E_{ik} + \beta s_{ik} + \eta_{ik} \quad (3)
\end{aligned}
$$

The first is a first stage predicting sewer percentage and using all three of our candidate instruments, $\mathbb{1}(\text{Outside})_{ik}$, $\mathbb{1}(\text{Outside})_{ik} x_{ik}$, and $\mathbb{1}(\text{Outside})_{ik} \Delta E_{ik}$. The second is the corresponding reduced form predicting a tract level outcome. Depending on estimation technique, the third is an OLS estimation yielding a correlation between sewer percentage and the outcome, or a TSLS regression yielding (subject to instrument validity) a causal estimate.

These equations require four comments. First, because our treatment is continuous, we assume homogeneous treatment effects. An estimate of $\beta$ in equation 3 is an average treatment effect by assumption.

Second, we have seen 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.

Table 2: Sewers and population density, Universe

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| *1. OLS* | | | | | | | | | |
| sewer % | 0.0086*** | 0.0086*** | 0.0086*** | 0.0082*** | 0.0082*** | 0.0082*** | 0.0074*** | 0.0074*** | 0.0074*** |
| | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| *2. First stage* | | | | | | | | | |
| outside | -0.7760** | | -2.9260*** | -1.3785*** | | -0.8456** | -0.9757*** | | -0.6454 |
| | (0.3409) | | (0.3559) | (0.3457) | | (0.4184) | (0.3555) | | (0.4449) |
| x*Outside | -0.0050*** | | | -0.0055*** | | | -0.0064*** | | |
| | (0.0004) | | | (0.0005) | | | (0.0006) | | |
| $\Delta$Elev*Outside | | -0.0773*** | -0.0385*** | | -0.0442*** | -0.0349*** | | -0.0301*** | -0.0221** |
| | | (0.0050) | (0.0069) | | (0.0070) | (0.0086) | | (0.0073) | (0.0093) |
| *3. Reduced form pop density* | | | | | | | | | |
| outside | 0.0335** | | 0.0427** | 0.0286* | | 0.0807*** | 0.0045 | | 0.0338 |
| | (0.0147) | | (0.0178) | (0.0158) | | (0.0209) | (0.0160) | | (0.0215) |
| x*Outside | -0.0001*** | | | -0.0003*** | | | -0.0004*** | | |
| | (0.0000) | | | (0.0000) | | | (0.0000) | | |
| $\Delta$Elev*Outside | | -0.0017*** | -0.0023*** | | -0.0016*** | -0.0025*** | | -0.0010*** | -0.0014*** |
| | | (0.0003) | (0.0004) | | (0.0004) | (0.0005) | | (0.0004) | (0.0005) |
| *4. IV* | | | | | | | | | |
| sewer % | 0.0211*** | 0.0221*** | 0.0143*** | 0.0394*** | 0.0364*** | 0.0241*** | 0.0594*** | 0.0324** | 0.0225** |
| | (0.0036) | (0.0031) | (0.0025) | (0.0045) | (0.0086) | (0.0073) | (0.0059) | (0.0127) | (0.0108) |
| F | 88.88 | 297.8 | 188.9 | 84.72 | 43.57 | 24.04 | 72.20 | 17.92 | 10.15 |
| $N$ | 53775 | 53775 | 53775 | 53775 | 53775 | 53775 | 53775 | 53775 | 53775 |
| elevation | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| x | Y | | | Y | | | Y | | |
| $\pi$ bins | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| seg×x | . | . | . | Y | Y | Y | . | . | . |
| $\pi$ bin×x | . | . | . | . | . | . | Y | Y | Y |

Note: Robust standard errors in parentheses.
Significance stars * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Third, our three instruments are an indicator for 'outside', this indicator interacted with the distance to the boundary, $x_{ik}$, and this indicator interacted with meters of climbing required to reach the basin divide, $\Delta E_{ik}$. We are concerned that lateral displacement has a direct effect on sewer percentage. Given this, in our preferred specifications we control for lateral displacement by including the interaction of the radial bin indicator with the perpendicular displacement, that is, $\mathbb{1}(k)x_{ik}$. This means that the effect of other variables are conditional on a radial bin specific slope and intercept. In the first stage and reduced form equations, the coefficients $\alpha_0$, $\alpha_1$ and $\gamma_0$ and $\gamma_1$ reflect sample mean changes in the level and trend of sewer percentage and outcome relative to radial bin specific means and trends. This means that we are parameterizing the basin divide breaks in sewer percentage and log population density that we see in figure 8(a) and (c) with a change in level and slope. To the extent that crossing the basin divide is costly, and the resulting decrease in sewer access affects population density, we expect to observe a step down and more rapid decrease as $x_{ik}$ crosses the basin divide at zero and continues to increase.

Fourth, we are also concerned that elevation has an independent effect on sewer percentage and outcome. Given this, we also control for elevation. This means that the coefficients $\alpha_2$ and $\gamma_2$ measure the difference in the effect of slope for parcels on the outside of the basin divide relative to the inside. To the extent that vertical rise to the basin divide is more costly outside the basin than inside, and recalling that $\Delta E_{ik} \geq 0$, we expect $\alpha_2$ and $\gamma_2$ to be negative. That is, the decline in sewer percentage and log population density should be more rapid as we move outside from the basin divide in figure 8 (b) and (d) then when we move inside.

Table 2 presents our main set of estimation results. The top panel gives results of OLS regressions of equation 3 using different controls. The second panel presents first stage regressions, equation (1) using different combinations of instruments and controls. The third panel is reduced form estimations, equation (2), mirroring the first stage estimations in the second panel, The fourth panel presents TSLS estimates of the effects sewer percentage on log population density using the instruments and specification common to other results in the same column. The bottom panel of the table describes the controls used in each specification, gives the sample size, and an F-statistic for the instruments used in the first stage regression of panel 2.

The columns are in three groups of three. The first three columns (1-3) present results for specifications that control for tract elevation, tract perpendicular distance to the basin divide, and radial bin intercepts The second group of three (4-6) allow the parameter on perpendicular distance to vary by basin divide segment, and the third group of three (7-9) allows this slope to vary by radial bin. The third group of three provides the most flexible control for potential confounding trends in $x$ and is our preferred set of specifications. In results presented later, we will sometimes rely on the other specifications when we are considering subsamples that are too small to allow us to estimate this many parameters precisely.

Within each group of three columns, we vary the instruments that we use. In the first column of each set (1,4,7) our instruments are the outside indicator and the outside indicator interacted with distance to the boundary. The coefficients on these two variables give the sample mean decline and rate of decline in sewer percentage or outcome. In the second column of each set (2,5,8) we use only the interaction of elevation to the basin divide and the outside indicator. The coefficient on this variable tells the amount by which sewer percentage or log population density decrease with each additional meter of climbing required to reach the basin divide. The third column of each set of three (3,6,9) is mirrors the column to its left, but also includes the outside indicator as an instrument. The interpretation of this coefficient is the same as for columns (1,4,7). Because the OLS regression presented in panel one do not include instrumental variables, the OLS specification is identical for each of the three columns within a group, e.g., columns (1-3).

In the top panel of table 2 we see that a 1% increase in sewer percentage is associated with about a 1% increase in population density. This effect is estimated precisely and is stable across specifications.

The second panel presents first stage results. We see that all point estimates for instruments have the expected negative signs in every specification. The coefficients on the two interaction variables are stable across specifications and are estimated precisely. In every specification. the relevant F-statistic is above the threshold for conventional weak instrument tests. The fact that these coefficients are negative broadly confirms our intuition about the cost structure for providing sewer service. Whether measured by horizontal or vertical displacement, being on the outside of central basin divide increases the cost of sewer access and decreases its prevalence.

The third panel reports reduced form results that parallel the first stage results of the second panel. These results are less tidy than the first stage results. The coefficients on the interaction terms are consistently negative and are estimated precisely. The coefficient on the step is generally positive, but its magnitude varies across specifications and it is not always estimated precisely. These results support the idea that moving outside the basin divide, whether measured by horizontal of vertical travel, decreases population density, even if the instantaneous effect of crossing the divide is ambiguous.

Finally, panel 4 presents out TSLS results. These results indicate an elasticity between 1.4 in column 3 and 5.9 in column 7. Focusing on our preferred specification in the last group (7-9), this range is from about 2.2 to 5.9. Comparing with the OLS elasticities in panel 1, this suggests that the causal effect of sewer access is between two and six times as large as the correlation in the raw data.

Appendix tables present country estimation results identical to those in table 2, but broken out for the countries of Colombia, Brazil and South Africa (sample sizes for Tanzania and Jordan are too small).

## 6 Estimating treatment effects from aggregate data

In the estimations above, we consider an econometric model with a continuous treatment, 'share of sewer access in a tract', and controls. It is well known that this makes the interpretation of the estimand as a LATE more complicated than in the canonical discrete instrument and no controls case.

To address this problem, note that at the parcel level, treatment is binary, a parcel either has sewer access or it does not. This opens the door to the possibility of estimating treatment effects at the (latent) parcel level from tract data aggregating treatment and outcomes to the tract level. This is exactly the strategy we pursue.

In particular, the MTE model is well suited to estimating treatment effects in instrumental variable estimations with controls. We show that it is possible to estimate this model from aggreagate data by using a small variance approximation to estimate a parcel level model from tract level data. Informally, if we observe two tracts with the same sewer share but different outcomes and different variances of control variables, then these tracts can teach us about the effect of treatment and controls at the parcel level. We pursue exactly this strategy to estimate tract level average treatment effects.

**TBD**

## 7 Discussion

Results so far establish an effect size. Adding 1% of sewer connections causes and increase in population density of about 3%. It is difficult to assess whether this is an economically important effect.

To develop some intuition around this issue, we conduct the following exercise. We define a 'city' as consisting of all census tracts inside, or within two kilometers of the central basin. For each city in our sample, consider a counterfactual case where we add 1% to the count of sewer connections in the city. 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.

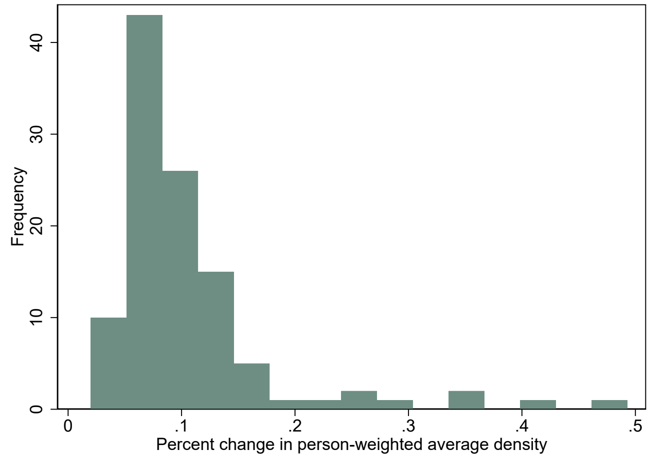
For each city, this process results in a counterfactual city where a subset of tracts has better sewer access than in the observed case. We can then use our estimates of treatment effects to calculate the implied increase in population in these tracts. Assuming that the new sewer connections increase city population by inducing rural residents to migrate to the city and that our treatment effect is 3% then, mechanically, our counterfactual cities house 3% more people than their observed counterparts.

This effects seems large in the following sense. Baum-Snow (2007) finds that each radial interstate highway decreased the density of US central cities by 9% over about a 40 year period. Our estimates suggest the opposite effect can be accomplished by adding about 3% to a central city's stock of sewer connections. That is, a 1% increment to a central city's sewer share is about one third as important for urban form as is a limited access radial highway.

The increase in person weighted density is also of intereest, but must be evaluated tract by tract. We perform this calculation for all 108 cities in our sample. Figure 9 presents a histogram summarizing our results. The modal city in our sample experiences an about 10% increase person weighted density, and the upper tail experiences much larger increases.

These effect also seems large. The relationship between density and labor productivity is well established, and a central estimate is that doubling the density a city increases wages by about 5%. Combining this estimate with the modal 10% increase in person weighted density we see in figure 9, suggests that adding 1% to the stock of sewer connections will increase wages of incumbent residents by about 0.5%. This amount is a flow. Taking its discounted present value using a 5% interest rate suggests that it is worth about 10% of the city's total annual wage bill. We suspect that this benefit alone will often be large relative to the cost of adding the required 1% of connections. Including the likely health benefits to incumbents, and the wage increase experienced by new

Figure 9: Histogram of counterfactual changes in density




Note: *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.*

migrants will increase this estimate of benefits further.

**A** ***CBD access and sewer access***

It is now common to evaluate public transit systems on the basis of the extent to which they improve access to the central city and thereby improve the functioning of the labor market. By facilitating higher residential densities, improving the sewer network within walking distance of the CBD, sewer networks can have the same effect.

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, walking distance. We find that effects are heterogeneous across cities, but that in many, mostly Latin American cities, these effects are large compared to those of a successful BRT system.

**TBD**

## 8 Conclusion

We estimate the effects of sewer access on population density in a sample of developing world cities using a novel identification strategy that derives from principles of wastewater engineering. Our estimates suggest that adding 1% is about one third as important for population density as is access to a limited access radial highway. This effect size implies that sewer connections likely have effects on the density of cities which lead to increases in labor productivity that are plausibly large compared to the construction costs of sewers.

## 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.

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.

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

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

Coury, M., Kitagawa, T., Shertzer, A., and Turner, M. (2022b). 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.

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

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

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 p acific and s outheast a sia. *Asia & the Pacific Policy Studies*, 1(1):18–32.

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.

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.

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.

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]. https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003. Last accessed 19 May 2022.

NASA/METI/AIST/Japan Spacesystems and U.S./Japan ASTER Science Team (2019). Aster global digital elevation model v003. https://doi.org/10.5067/ASTER/ASTGTM.003. 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.

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

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.

Table 3: Estimation sample; Brazil, Colombia, South Africa, Jordan

| Country | cities | $\pi$-bins | tracts | share inside | Tract area km$^2$ | People/km$^2$ | Sewer percentage |
|---|---|---|---|---|---|---|---|
| Brazil | 62 | 2164 | 40,117 | 0.45 | 1.27 | 15,603 | 0.68 |
| Colombia | 18 | 1049 | 36,706 | 0.51 | 0.47 | 28,667 | 0.70 |
| Jordan | 6 | 75 | 172 | 0.46 | 5.32 | 871 | 0.26 |
| South Africa | 15 | 499 | 7,927 | 0.52 | 1.95 | 8,226 | 0.76 |
| Tanzania | 7 | 387 | 3918 | 0.68 | 2.09 | 8,677 | 0.01 |

Note:*Jordan is households, not people. The economic geography of these places is really different. Estimations will be at tract/bin level, so most of the weight will come from Colombia and Brazil. Tanzania has no sewers.*

# Appendix A Supplemental results

Table 4: Sewers and population density, Brazil

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| *1. OLS* | | | | | | | | | |
| sewer % | 0.0079$^{***}$ | 0.0079$^{***}$ | 0.0079$^{***}$ | 0.0074$^{***}$ | 0.0074$^{***}$ | 0.0074$^{***}$ | 0.0068$^{***}$ | 0.0068$^{***}$ | 0.0068$^{***}$ |
| | (0.0004) | (0.0004) | (0.0004) | (0.0004) | (0.0004) | (0.0004) | (0.0004) | (0.0004) | (0.0004) |
| $N$ | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 |
| *2. First stage* | | | | | | | | | |
| outside | -0.3898 | | -4.8475$^{***}$ | -0.9176$^{*}$ | | -1.7200$^{***}$ | -1.0832$^{**}$ | | -1.8099$^{***}$ |
| | (0.4740) | | (0.4623) | (0.4875) | | (0.5697) | (0.4969) | | (0.6008) |
| x*Outside | -0.0019$^{***}$ | | | -0.0034$^{***}$ | | | -0.0054$^{***}$ | | |
| | (0.0006) | | | (0.0007) | | | (0.0008) | | |
| $\Delta$Elev*Outside | | -0.0430$^{***}$ | 0.0153$^{**}$ | | -0.0047 | 0.0124 | | -0.0035 | 0.0163 |
| | | (0.0054) | (0.0069) | | (0.0079) | (0.0096) | | (0.0081) | (0.0103) |
| $N$ | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 |
| *3. Reduced form pop density* | | | | | | | | | |
| outside | -0.0172 | | -0.0931$^{***}$ | -0.0361$^{*}$ | | -0.0163 | -0.0693$^{***}$ | | -0.0485$^{*}$ |
| | (0.0205) | | (0.0201) | (0.0209) | | (0.0245) | (0.0214) | | (0.0257) |
| x*Outside | -0.0001$^{***}$ | | | -0.0002$^{***}$ | | | -0.0004$^{***}$ | | |
| | (0.0000) | | | (0.0000) | | | (0.0000) | | |
| $\Delta$Elev*Outside | | -0.0021$^{***}$ | -0.0010$^{**}$ | | -0.0014$^{***}$ | -0.0012$^{**}$ | | -0.0017$^{***}$ | -0.0011$^{**}$ |
| | | (0.0003) | (0.0004) | | (0.0004) | (0.0005) | | (0.0004) | (0.0005) |
| $N$ | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 |
| *4. IV* | | | | | | | | | |
| sewer % | 0.0488$^{***}$ | 0.0496$^{***}$ | 0.0281$^{***}$ | 0.0691$^{***}$ | 0.2996 | 0.0169 | 0.0694$^{***}$ | 0.4749 | 0.0325$^{**}$ |
| | (0.0163) | (0.0078) | (0.0038) | (0.0135) | (0.4903) | (0.0140) | (0.0106) | (1.0772) | (0.0161) |
| $N$ | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 |
| F | 6.885 | 51.46 | 88.66 | 17.43 | 0.259 | 5.081 | 28.97 | 0.132 | 5.227 |
| $N$ | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 | 27373 |
| elevation | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| x | Y | | | Y | | | Y | | |
| $\pi$ bins | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| seg$\times$x | . | . | . | Y | Y | Y | . | . | . |
| $\pi$ bin$\times$x | . | . | . | . | . | . | Y | Y | Y |

Note: Robust standard errors in parentheses.

Significance stars * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Sewers and population density, Colombia

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| *1. OLS* | | | | | | | | | |
| sewer % | 0.0112$^{***}$ | 0.0112$^{***}$ | 0.0112$^{***}$ | 0.0107$^{***}$ | 0.0107$^{***}$ | 0.0107$^{***}$ | 0.0097$^{***}$ | 0.0097$^{***}$ | 0.0097$^{***}$ |
| | (0.0004) | (0.0004) | (0.0004) | (0.0004) | (0.0004) | (0.0004) | (0.0004) | (0.0004) | (0.0004) |
| $N$ | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 |
| *2. First stage* | | | | | | | | | |
| outside | -0.3793 | | 2.4720$^{***}$ | -1.1387$^{**}$ | | 1.9094$^{***}$ | -0.4779 | | 2.5140$^{***}$ |
| | (0.5131) | | (0.7183) | (0.5134) | | (0.6834) | (0.5361) | | (0.7262) |
| x*Outside | -0.0100$^{***}$ | | | -0.0100$^{***}$ | | | -0.0095$^{***}$ | | |
| | (0.0006) | | | (0.0007) | | | (0.0008) | | |
| $\Delta$Elev*Outside | | -0.1157$^{***}$ | -0.1527$^{***}$ | | -0.0927$^{***}$ | -0.1171$^{***}$ | | -0.0656$^{***}$ | -0.1029$^{***}$ |
| | | (0.0098) | (0.0177) | | (0.0131) | (0.0171) | | (0.0138) | (0.0185) |
| $N$ | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 |
| *3. Reduced form pop density* | | | | | | | | | |
| outside | 0.1044$^{***}$ | | 0.3023$^{***}$ | 0.1162$^{***}$ | | 0.2862$^{***}$ | 0.1014$^{***}$ | | 0.2149$^{***}$ |
| | (0.0222) | | (0.0407) | (0.0254) | | (0.0418) | (0.0252) | | (0.0433) |
| x*Outside | -0.0001$^{***}$ | | | -0.0003$^{***}$ | | | -0.0004$^{***}$ | | |
| | (0.0000) | | | (0.0000) | | | (0.0001) | | |
| $\Delta$Elev*Outside | | -0.0015$^{***}$ | -0.0060$^{***}$ | | -0.0026$^{***}$ | -0.0062$^{***}$ | | -0.0007 | -0.0039$^{***}$ |
| | | (0.0005) | (0.0010) | | (0.0007) | (0.0011) | | (0.0007) | (0.0011) |
| $N$ | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 |
| *4. IV* | | | | | | | | | |
| sewer % | 0.0136$^{***}$ | 0.0126$^{***}$ | 0.0203$^{***}$ | 0.0229$^{***}$ | 0.0277$^{***}$ | 0.0399$^{***}$ | 0.0429$^{***}$ | 0.0113 | 0.0320$^{***}$ |
| | (0.0031) | (0.0038) | (0.0039) | (0.0039) | (0.0072) | (0.0076) | (0.0055) | (0.0100) | (0.0098) |
| $N$ | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 | 23199 |
| F | 141.2 | 272.3 | 146.5 | 106.8 | 84.73 | 47.06 | 65.91 | 35.22 | 24.45 |
| elevation | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| x | Y | | | Y | | | Y | | |
| $\pi$ bins | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| seg$\times$x | . | . | . | Y | Y | Y | . | . | . |
| $\pi$ bin$\times$x | . | . | . | . | . | . | Y | Y | Y |

Note: Robust standard errors in parentheses.
Significance stars * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Sewers and population density, South Africa

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| *1. OLS* | | | | | | | | | |
| sewer % | -0.0008 | -0.0008 | -0.0008 | 0.0000 | 0.0000 | 0.0000 | -0.0017$^{*}$ | -0.0017$^{*}$ | -0.0017$^{*}$ |
| | (0.0009) | (0.0009) | (0.0009) | (0.0010) | (0.0010) | (0.0010) | (0.0010) | (0.0010) | (0.0010) |
| *N* | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 |
| *2. First stage* | | | | | | | | | |
| outside | -6.3020$^{***}$ | | -10.0796$^{***}$ | -7.0174$^{***}$ | | -5.3771$^{***}$ | -3.9864$^{**}$ | | -3.5805$^{*}$ |
| | (1.6134) | | (1.3776) | (1.6123) | | (1.6787) | (1.6491) | | (1.8895) |
| x*Outside | -0.0000 | | | 0.0013 | | | 0.0065$^{**}$ | | |
| | (0.0019) | | | (0.0022) | | | (0.0029) | | |
| $\Delta$Elev*Outside | | -0.1890$^{***}$ | -0.0445 | | -0.1377$^{***}$ | -0.0772$^{*}$ | | -0.0977$^{**}$ | -0.0512 |
| | | (0.0238) | (0.0274) | | (0.0366) | (0.0417) | | (0.0384) | (0.0478) |
| *N* | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 |
| *3. Reduced form pop density* | | | | | | | | | |
| outside | -0.0219 | | -0.1290$^{*}$ | -0.0120 | | -0.1465$^{*}$ | -0.0334 | | -0.2137$^{**}$ |
| | (0.0700) | | (0.0697) | (0.0716) | | (0.0811) | (0.0801) | | (0.0908) |
| x*Outside | -0.0000 | | | -0.0002$^{*}$ | | | -0.0005$^{***}$ | | |
| | (0.0001) | | | (0.0001) | | | (0.0001) | | |
| $\Delta$Elev*Outside | | 0.0019$^{*}$ | 0.0037$^{**}$ | | 0.0046$^{**}$ | 0.0063$^{***}$ | | 0.0070$^{***}$ | 0.0097$^{***}$ |
| | | (0.0011) | (0.0015) | | (0.0018) | (0.0022) | | (0.0020) | (0.0025) |
| *N* | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 |
| *4. IV* | | | | | | | | | |
| sewer % | 0.0034 | -0.0099$^{*}$ | -0.0004 | -0.0028 | -0.0337$^{**}$ | -0.0087 | -0.0313$^{**}$ | -0.0712$^{**}$ | -0.0201 |
| | (0.0112) | (0.0058) | (0.0045) | (0.0101) | (0.0165) | (0.0095) | (0.0159) | (0.0341) | (0.0160) |
| *N* | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 | 3177 |
| F | 7.539 | 64.73 | 56.57 | 9.524 | 13.31 | 11.32 | 6.383 | 5.806 | 4.766 |
| elevation | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| x | Y | | | Y | | | Y | | |
| $\pi$ bins | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| seg$\times$x | . | . | . | Y | Y | Y | . | . | . |
| $\pi$ bin$\times$x | . | . | . | . | . | . | Y | Y | Y |

Note: Robust standard errors in parentheses.
Significance stars * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

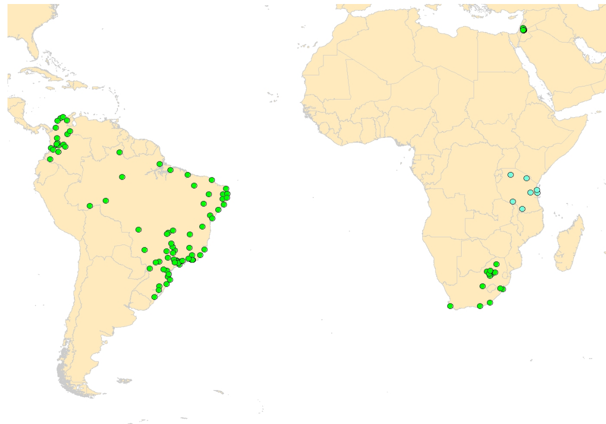
# Appendix B Data

## *A. Cities*

The UN DESA World Urbanization Prospects data (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 in the following subsections, we estimate treatment effects using all cities in the UN Cities data in Brazil, Colombia, South Africa, and Jordan. We can also evaluate counterfactuals in

Figure 10: Locations from UN Cities data in our sample



Note: *Data from the UN DESA World Urbanization Prospects (DESA Population Division, 2018), which is a census of all cities that had a population 300,000 or more in 2014. We select cities in Brazil, Colombia, South Africa, and Jordan for estimation. Tanzania, colored differently above, is only included in the counterfactual exercise.*

these cities as well as those in Tanzania. A map of the resulting cities is displayed in Figure 10, where Tanzanian cities are colored differently to indicate their inclusion only in the counterfactual exercises.

***B. Sewers***

Each of the countries we study provide comprehensive surveys of sewer access at granular geographies in the 2010s. We calculate the percentage 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 exists between a census "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?"

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?"

### *C. Population density, other outcomes*

Using the same censuses described in the previous subsection, we are able to calculate population density and literacy measures for all countries, as well as income measures for Brazil and South Africa. 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 at the monthly level in both Brazil and South Africa. Brazil reports the average nominal monthly income of head of household at the setores-level. 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 SAL area by assining each income bin the midpoint of the income range for the same bin, then using the reported count to calculate the mean. Income is not reported in Colombia, Tanzania, or Jordan 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.

***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 U.S./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 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 Uberlandia, Brazil. This is an empirical analog of the basins drawn in figure 2. In figures 3, black lines indicate the boundaries of drainage basins. 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. 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. SRTM is about the same. (Uuemaa et al., 2020).