How much more expensive is housing in larger cities? Worldwide evidence from Airbnb[1]

Bernhard Nöbauer[2]

University of Lausanne

September 2023

Abstract

Using a hedonic regression approach with data from 1.53 million Airbnb properties, I estimate the price of a representative short-term rental property at the center of 734 cities worldwide. The estimated rental prices provide an internationally standardized proxy for housing costs. Rental prices computed in this way are found to be highest in Amsterdam, London, New York, and San Francisco. I use these standardized rental price estimates to compute the elasticity of housing costs with respect to city size. My preferred specification shows an elasticity of 0.16, statistically significant at the 1% level. However, there is considerable geographic heterogeneity. Housing costs increase more strongly in city size in the euro area and India than elsewhere. In contrast, I find them to decrease in city size in Mexico. I offer suggestive evidence that crime might explain this unusual result.

[1] I thank Marius Brülhart for his guidance and countless hours of fruitful discussion. Moreover, I thank Dzhamilya Nigmatulina for helpful comments and Gilles Duranton for his inspiration. Moreover, I thank Laura Camarero Wislocka for her excellent help with assessing and defining city centers.

[2] Departement of Economics, Faculty of Business and Economics (HEC Lausanne), University Of Lausanne, 1015 Lausanne, Switzerland; `bernhard.noebauer@bluewin.ch`.

1 Introduction

A lot of empirical work confirms that wages are increasing in city size (for a survey see, for example, Combes and Gobillon, 2015). Bigger cities offer better opportunities to learn, share, and match, forces that are commonly summarized under agglomeration economies (Duranton and Puga, 2004). However, since our societies have not converged to live in a single gigantic city, there must be costs that make big cities less efficient or pleasant and that at least partially counteract agglomeration economies (Henderson, 1974). A particularly prominent example of such costs is housing costs, which are the focus of this paper.

The elasticity of housing costs with respect to city size measures how much more expensive housing becomes when city size increases, and estimates of that elasticity are surprisingly scarce. A prominent exception is Combes et al. (2018), who measure an elasticity of house prices with respect to city size of 0.21 for a sample of 277 urban areas in France. Given the French context, their estimates are mainly based on mid-sized cities, with an average urban area population of 166,020 and a median of 47,909 (p. 1565). However, we might expect urban costs to be disproportionally higher in the largest cities. Combes et al. (2018) show evidence of that by estimating a non-linear effect of city size, but these estimates are based on few observations at the upper end of the French city size distribution. I supplement their evidence using a worldwide sample of 734 cities with an average population of 2,100,936 and a median of 905,270 that has more to say about the housing cost premium of large and very large cities and goes beyond the context of a developed country.

Methodologically, I follow Combes et al. (2018) in measuring housing costs at the city center. This has the advantage that differences in transportation costs have a smaller influence on comparisons across cities, or no influence at all if we take the monocentric city model at face value. This model still guides a lot of research in urban economics and its assumptions are widely applied (for a survey on the model and its application see Duranton and Puga, 2015). In contrast, comparing city average real estate prices comes with the problem that the average property in a big city like Tokyo is further away from the center than the average property in a smaller city like Kagoshima, which implies higher transportation costs that confound the comparison.

For my analysis, I use data on short-term rental properties from Airbnb. Using these novel data allows me to extend the analysis to the global scale, based on an extensive set of variables that describe the properties in an internationally standardized way. This worldwide scope is hard to achieve with traditional data from national statistical offices or real estate platforms. As Airbnb hosts typically compete for the same housing units as long-term residents, across-city differences in nightly rates serve as a proxy for differences in long-term housing costs. I show, for the examples of France and the United States, that the city comparisons of housing costs estimated with Airbnb properties correspond to those estimated with long-term rental objects, albeit not perfectly.

I choose the 734 cities in the sample, their geographic boundaries, and their center points using transparent rules that I apply worldwide. Within these cities, I have data on 1.53 million properties that were active between January 2018 and March 2019 and available or rented for at least 100 out of 365 days. I do this sample restriction to exclude apartments that only capitalize on peak price periods, using the Men's Fifa World Cup in June and July 2018 in Russia as a natural experiment to determine the cutoff value.

Using a hedonic regression with city fixed effects and city-specific distance gradients, I create a ranking of the 734 cities regarding the rental rate of a representative property at the city center. That property can be rented in its entirety by a maximum of two guests, has one bedroom and one bathroom, and shares the standard of an average Airbnb property in its city regarding all other characteristics. Given my methodology and data, I estimate Amsterdam, London, New York, and San Francisco to have the highest rental prices in the world. Caracas, Mandalay, Monteria, and Srinagar are at the other end of the ranking, with rental prices that are around 20 times lower. I run multiple robustness checks to confirm that the ranking is robust to changing underlying assumptions.

In the second stage, I regress the estimated rental prices on city size. For most specifications, I follow Combes et al. (2018) in using log population to measure city size and controlling for log area. This setup can be read as an unrestricted version of population density. I include country fixed effects when using the worldwide sample, so the coefficients are estimated from within-country variation. Moreover, I control for various city characteristics, including for the number of Airbnb properties per 100,000 inhabitants to control for the attractiveness of a city to tourists. An instrumental variable approach in which historical population sizes are used as an instrument serves as a robustness check. In my preferred specification, I estimate an elasticity of 0.161. This coefficient implies that a 10% higher population size is associated with housing costs that are 1.61% higher. The effect is statistically significant at the 1% level.

The literature provides a small number of related results. Ahlfeldt and Pietrostefani (2019) suggest an elasticity of rent with respect to population density of 0.15,[1] while Henderson (2002) estimates an elasticity of the rent to income ratio with respect to metro area size of 0.32. When not controlling for area, my results for the elasticity of housing costs with respect to city size almost precisely match those of Combes et al. (2018). They report an elasticity of 0.11, while my estimation yields an elasticity of 0.12. Combes et al. (2018) interpret this specification as the costs of unrestricted city size, while controlling for area corresponds to a city that is restricted from expanding outwards. The estimates suggest that the costs of unrestricted city size are very similar in our two contexts, with the difference between our main estimates coming exclusively from the area-restricted version. An intuitive explanation for this finding could the be stringent building height regulations in France (Jedwab et al., 2022). If a city is not allowed to expand outwards, constructing higher buildings is one of the remaining solutions to accommodate a larger population. The extent to which this

[1] Ahlfeldt and Pietrostefani (2019) is a meta-study that discusses the effects of density on multiple outcome variables. When I use density, instead of the more flexible specification of population and area, I obtain an estimate of 0.21, which is statistically significant at the 1% level.

solution is embraced will affect the increase in housing costs associated with a growing population.[2]

My work also expands the evidence on the geographical heterogeneity of the elasticity of housing costs with respect to city size. I compute separate regressions for the six countries with the highest number of cities in the sample (the United States, Russia, China, India, Brazil, and Mexico) and for the eurozone. The estimated elasticity is above the global average for the United States and Russia and is particularly high in the eurozone and India.[3] The estimate for the eurozone is within 0.04 percentage points from what Combes et al. (2018) estimate for comparably large French cities when using their non-linear specification. These findings point again towards an above average elasticity of housing costs with respect to city size for large European cities. While Chauvin et al. (2017) focus on agglomeration economies rather than urban costs, their work includes estimations of the elasticity of housing costs with respect to city size for the United States, Brazil, China, and India. My results are similar to theirs for the US, China, and Brazil, but they are very different for India, where Chauvin et al. (2017) do not find any effect of city size on housing rents. However, they do control for neither property nor city characteristics and they estimate the price of an average housing unit instead of a housing unit at the city center. When I apply their second-stage estimation strategy, I also find an elasticity that is indistinguishable from zero.

Being surrounded by many people might not always be beneficial. For Mexico, I estimate a statistically significantly negative coefficient for population and a statistically significantly positive coefficient for area. This finding implies that denser cities are cheaper in the Mexican context. I conjecture that crime might be a driver of this finding. The country is in the midst of a drug war (see, for example, Shirk and Wallman, 2015) and safety concerns are probably more important than elsewhere. I explore this hypothesis by adding an interaction term between log population and a city's homicide rate. My results show that Mexican cities with high homicide rates have a statistically significantly more negative elasticity of housing costs with respect to city size. I then proceed with the global sample and test the interaction between city size and an indicator for being among the 50 cities with the world's highest homicide rates. Cities in that group have a less positive elasticity of housing costs with respect to city size. The difference is statistically significant and quantitatively large, with an estimated elasticity that is more than 40% lower. The finding that crime lowers the elasticity of housing costs with respect to city size complements the evidence that large cities are more affected by crime (Glaeser and Sacerdote, 1999) and that crime negatively affects house prices (Pope and Pope, 2012).

The remainder of the paper is organized as follows: Section 2 describes the main data and discusses city definitions. Section 3 presents the first-stage hedonic regression and results in a ranking of cities by their estimated short-term rental price at the city center. In Section 4, I use these rental prices as an input for the second-stage regressions, and I present and discuss the corresponding

[2] French central cities often have many beautiful old buildings, and tearing them down is probably not an optimal solution. Glaeser (2011) discusses this nexus and possible ways forward using the example of Paris.

[3] The coefficients are statistically significant at the 5% level for the US, India, and the eurozone, but not for Russia.

estimation results. Section 5 concludes. Finally, the appendix contains the full city ranking and several auxiliary results.

2 Data and city definitions

This project relies on two main types of data: geolocalized data on Airbnb properties and spatially disaggregated population data. I combine the latter with data on city centers and the boundaries of urban areas to define 734 cities that I include in this study. All of these data are available in an internationally standardized way, which allows me to conduct the analysis on a global scale. This section will successively present the data on the Airbnb properties, the city definitions I use and the data on population sizes of the resulting cities.

Airbnb properties

The data on short-term rental properties from Airbnb come from AirDNA, a company specialized in "short-term rental data and analytics".[4] They contain close to all properties that were advertised on Airbnb at least once between 2018-01-01 and 2019-03-25.[5] By combining information about days for which properties are rented with information about the price for these days, AirDNA is able to estimate the prices actually paid by customers. For every property, I have information about the average daily price over the twelve months before the date on which a property was last web scraped from the Airbnb website. I also have the coordinates of the location for each property, even though some of them are scrambled within a short radius due to security concerns.[6] Moreover, the data contain a substantial number of covariates, from the number of bedrooms to the presence of a hairdryer. All of these variables are available in an internationally standardized way. Overall, I can match 3.07 million properties to the 734 cities in my sample, 1.53 million of which were available for rent or rented for at least 100 of the last 365 days before they were last scraped.[7]

Cities

It is not straightforward to find a definition of where a city ends and where its center is located. The problem becomes especially complicated if the definition is supposed to work well for very different

[4] `https://airdna.co`, last accessed: 2023-01-16.

[5] To the best of my knowledge, AirDNA web scraped every single property from Airbnb once every three days over this period. This implies that a small number of properties that appeared only briefly and were immediately removed or rented might not be part of the dataset.

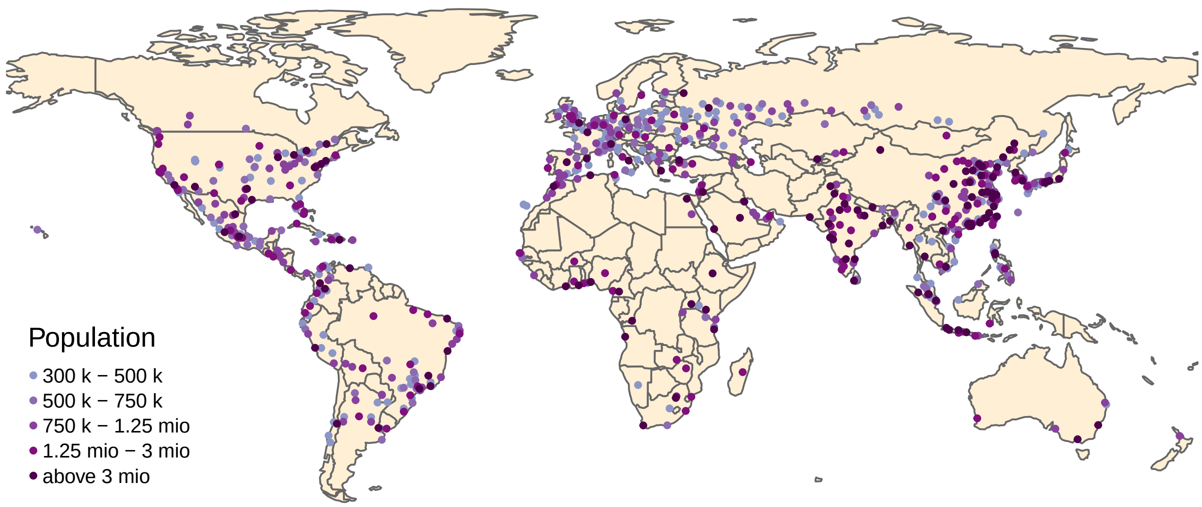
[6] Airbnb recommends that hosts indicate their precise address. However, hosts are free to choose whether they prefer a precise pin to be shown at the address of their property, or a circle that indicates the approximate location in a close radius (`https://www.airbnb.com/resources/hosting-homes/a/setting-expectations-with-an-accurate-location-491?_set_bev_on_new_domain=1686948204_M2FkMTUxYWQzNzNi&locale=en`, last accessed: 2023-06-16). The current maximal deviation from the true location of the property is indicated as 800 meters. When I obtained the data in 2019, AirDNA suggested an even smaller maximal deviation of 500 meters.

[7] The raw dataset contains 9,419,495 observations. However, 2,354,445 of these properties were never reserved. I have to drop another 1,917 observations because their coordinates are missing. Afterwards, I can spatially join 3,093,755 properties with my city polygons. An additional 25,603 observations drop out because of missing covariates (or a missing price in one case). In the end, 3,068,152 entries remain.

countries. Here, I explain the main decisions I make to come to a definition that I deem suitable for the empirical exercise I conduct.[8]

I start with the open collaboration database platform OpenStreetMap that relies on crowd intelligence. This website asks users to place so-called city tags "at the center of the city, like the central square, a central administrative or religious building or a central road junction".[9] These geolocalized tags are my first candidates for both cities' locations and city centers. As a second step, I spatially join the city tags to all urban center polygons of the Global Human Settlement Layer project (Florczyk et al., 2019) with a population size of at least 300,000.[10] In some cases, urban center polygons contain multiple city tags. Using the city population counts from OpenStreetMap, I retain all city tags that are associated with a population count that amounts to at least 40% of the highest population count in an urban center.[11]

Figure 1: 734 cities in the sample



Note: The dots in this figure show the geographic distribution of the 734 cities in my sample. Their colors refer to the city's population size, with larger cities represented in darker shades. To be included in the sample, a city must have had a population of at least 300,000 inhabitants in 2015 and at least 100 Airbnb properties that were active between January 2018 and March 2019.

I then manually check all remaining tags using satellite and street view images from GoogleMaps

[8] For more details, refer to Nöbauer (2023), where I use the same city definitions and delineations and describe them more extensively.

[9] `https://wiki.openstreetmap.org/wiki/Tag:place%3Dcity`, last accessed: 2023-01-13. The website also includes information about many different kinds of geographic tags like motorways, restaurants, or playgrounds. As of 2023-06-12, the website contains 2.951 billion tags in total (`https://taginfo.openstreetmap.org/reports/database_statistics`).

[10] The basis of the urban centers of Florczyk et al. (2019) are contiguous 1km × 1km grid cells with an estimated population of at least 1,500 or a built-up area of at least 50%. Their definition results in some cities being very broad and containing multiple well-known cities, for example Oakland/San Francisco/San José or Kobe/Kyoto/Osaka. In some of these cases setting one city center for the whole urban area would be very tricky. I therefore decide against simply adopting their definition. I believe that combining the urban centers from Florczyk et al. (2019) with data on cities from OpenStreetMap results in a set of cities that is better suited for this analysis.

[11] I complement missing population counts on OpenStreetMap with information from Wikipedia.

and assess whether they are an appropriate choice for the city center. Whenever this is not the case, I resort to center coordinates from Google Maps. If they also describe a point that visually does not constitute a suitable center, I provide my own best guess. Once the city centers are determined, I split urban centers with multiple remaining city tags that are more than 7 km apart, so that it is likely that they constitute two distinct cities.[12] In a final step, I recompute the population size and the number of Airbnb properties within each city and retain those with at least 300,000 inhabitants and 100 properties.[13] Figure 1 shows the geographic distribution of the resulting 734 cities.

Population counts

The population counts also come from the Global Human Settlement project, more precisely from the GHS-POP file (Schiavina et al., 2019). They take population data from administrative sources at the smallest availbale scale. They then disaggregate these data to 1 km $\times$ 1 km grid cells using the proportion of buildings and other artificial structures, detected from day-light satellite images, with machine-learning.[14] They apply the same procedure to satellite images from 2015 and 1975, which ensures a certain level of intertemporal comparability that is beneficial to my instrumental variable approach.[15]

3 First-stage regressions

When assessing the effect of city size on real-estate prices, it is important to use housing units that are as comparable as possible across cities. Even with standardized data from a single source, simply computing the average price of units for each city is insufficient. There are two main reasons for this: First, the size and quality of housing units vary non-randomly. For example, the average apartment in Paris has fewer bedrooms than the average apartment in Toulon, while the average one-bedroom apartment in Ho Chi Minh City has more amenities than the average one-bedroom apartment in Can Tho. Second, differences in the geographical expanse of cities imply differences in accessibility and transportation costs. The average apartment in Buenos Aires is much farther away from the city center than the average apartment in Salta, and its inhabitants might spend considerably more time commuting for work and leisure activities than their counterparts in Salta.

To address the first issue, I estimate a hedonic regression. Apart from a separate intercept for each

[12] To split the cities, I use a rule described by Akbar et al. (2021): Border points X are assigned such that $\text{dist}(X, A)/\text{dist}(X, B) = (\text{Pop A}/\text{Pop B})^{\frac{0.57}{2}}$ where $\text{dist}(X, A)$ denotes the distance of a grid point X to the center of city A and $\text{dist}(X, B)$ denotes the distance to the center of city B. I reassign enclaves in repeated iterations until there are no city parts left that do not contain a center. I split cities that span across two countries at the border, without reassigning enclaves.

[13] To assess the population size I use the population data presented below. As discribed above, the cutoff of 100 Airbnbs refers to properties that have been rented at least once.

[14] They also offer a 100 m $\times$ 100 m resolution. However, I keep the 1 km $\times$ 1 km grid structure of the GHSL urban centers for my cities, so I would not gain anything by using the better resolution.

[15] As discussed in Section 4.1, 40 years are hardly enough for a credible identification of the instrument. Nevertheless, it is progress to have intertemporally comparable population data on a global scale. I therefore present the results of an IV specficiation, while cautioning that they should not be interpreted as more than a robustness check.

city, which is the variable of interest in this first stage, I control for numerous characteristics for each property, particularly for the number of bedrooms, bathrooms, and the maximum number of guests allowed.[16] I also include indicators for whether guests have the entire apartment for themselves or have to share the apartment or even their room. As the effect of these core characteristics may well be nonlinear, I allow for a flexible functional form by including them as categorical variables. The left column of Figure A1 shows the respective categories and their distribution in the data.

The same column also shows the distribution of other variables for which I control. The number of photos serves as a proxy for how much effort is put into creating the profile on Airbnb. For this variable, I also include a squared term, as I expect the marginal effect of additional photos to be diminishing and potentially even negative at a very high number of pictures. The number of properties a host has on the platform controls for the fact that certain hosts offer multiple properties. The final row of Figure A1 shows an indicator for whether a property is within 500 meters of an ocean or big lake.[17] Furthermore, I control for 43 amenities, examples of which include the presence of a tv, a hairdryer, or a first aid kit, as well as the availability of breakfast or free parking. Figure A2 displays the list of amenities, with the fraction of properties in which they are available in brackets.[18]

The variables in this second group are included either as indicators or modeled using a linear or quadratic functional form. Moreover, I demean them within each city. To see why this improves the estimation, consider the amenity “heating”. Without demeaning, there is a selection effect. Most properties without heating are located closer to the equator. They are not necessarily cheaper because of the lack of heating, which is unnecessary in the warmest climate zones. However, they are often located in countries with lower overall price levels.[19] Including this variable without demeaning would therefore result in an overestimation of the effect of heating by absorbing part of the city-fixed effects. As this first-stage regression aims to estimate the city-fixed effects as precisely as possible, demeaning helps avoid these biases.

To address the second issue, I follow Combes et al. (2018) in estimating the price of a property at the city center rather than the price of an average property in a city. The economic intuition

[16] Unfortunately, I do not have data about the square meter size of an apartment. However, customers usually do not have access to this information either. It is only available to them if the host explicitly puts it in the property description or if they have stayed there before. In all other cases, customers cannot consider it for their decision-making, and I, therefore, expect its influence on prices to be limited.

[17] This is the only variable that is not directly visible on Airbnb. Instead, the customers can infer it from a map provided on the website, although Airbnb sometimes scrambles the coordinates to some limited extent (500m at the very most) for security concerns. Moreover, the hosts seem to have a clear incentive to indicate a location close to a coast or beach in the description and the photos. To construct these indicators, I measure the air-line distance from a property to the closest ocean, sea, or big lake (at least 80km^2). To determine the location of waters, I use ESRI’s “World Water Bodies” layer (`https://arcgis.com/home/item.html?id=e750071279bf450cbd510454a80f2e63`, downloaded on 2023-10-10) and the HydroLAKES data from `https://hydrosheds.org/products/hydrolakes` (downloaded on 2023-01-01).

[18] I only include amenities that are present in at least 1% of properties. The data include another 34 amenities available in very few apartments.

[19] To some extent, this can also be the case within an individual country. For example, in Italy, heating will be more of a necessity in the northern part of the country, which is also the wealthier part of the country.

for this builds on two of the most well-known models in urban economics: The Rosen-Roback model (Rosen, 1979; Roback, 1982), which describes choices between cities, and the monocentric city model (Alonso, 1964; Mills, 1967; Muth, 1969), which describes choices within cities. The monocentric city model features households that work in the city center for a given wage, bear transportation costs for their commute, consume housing, and a composite good. In equilibrium, the unit cost of housing is more expensive closer to the city center, as people are willing to pay higher prices to avoid commuting costs. Ex-ante homogeneous agents can end up with different bundles of a location, housing consumption, and the composite good, with all bundles yielding the same utility. As long as this equalized within-city utility is given, it does not matter which bundle is taken for the comparison of agents across cities. It is convenient to make the comparison in the city center, where transportation costs are zero according to the model's assumptions. This choice, in turn, facilitates the comparisons between cities that underpin the utility equalization across cities in the Rosen-Roback model. In the model's equilibrium, wage differences and amenities counterbalance differences in housing costs. Measuring the housing costs at the city center implies that transportation costs can be left out of the comparison.[20] While the present paper is exclusively concerned with estimating the housing cost aspect of this comparison, it is important to bear this bigger picture in mind.

Empirically, I implement the measurement at the city center by estimating both an intercept μ_c and a distance gradient β_c for each city c. I add $+1$ to the distance to the city center to be able to interpret a distance of $\ln(1) = 0$ as the city center. My preferred first-stage regression has the form

$$\ln(\text{price})_{ic} = \mu_c + \beta_c \ln(\text{distance+1})_{ic} + \boldsymbol{\gamma}\mathbf{X}_{ic} + \boldsymbol{\delta}(\mathbf{Z}_{ic} - \bar{\mathbf{Z}}_c) + \varepsilon_{ic} \ , \tag{1}$$

where X_{ic} denotes a set of core categorical variables for the type of the listing and the number of bedrooms, bathrooms, and the maximum number of guests allowed. The baseline categories are the respective modes (see Table A1). $\mathbf{Z}_{ic}$ denotes the second set of variables and amenities that are included with a specified functional form and demeaned by city.

A city fixed effect μ_c, therefore, has the interpretation of the log price of an apartment in city c that is located at the city center, rented out in its entirety to a maximal number of two guests, has one bedroom and one bathroom, and characteristics that match the city average for all variables in $\mathbf{Z}_{ic}$. Taking the exponential of the city fixed effect yields the USD price of this representative property.

3.1 Excluding apartments available only during price spikes

The founders of Airbnb got the idea for their business when participants of a conference in San Francisco struggled to find available hotel rooms (Gallagher, 2017). Some long-term tenants also

[20] In reality, people face transportation costs even if they live at the very center of a city, and these transportation costs vary across cities. However, I consider it probable that the comparison at the city center minimizes both the level of transportation costs and the differences in transportation costs between cities.

rent out their apartments while they are on vacation. In these cases, Airbnb can contribute to a more efficient capacity utilization of living space rather than merely displacing one kind of occupant with another. However, these cases also threaten the validity of using Airbnb data for my study. Major events can lead to a temporary surge in price. Examples include events with changing venues, such as the Super Bowl, but also annually recurring events like Art Basel. One might argue that such events contribute to the general attractiveness of a city and that an increase in price is, therefore, justified. However, if there is also a surge in the properties offered on Airbnb to take advantage of the temporarily higher prices, these marginal properties will have an average nightly rate that is much higher than what could be charged on a yearly basis. These properties will therefore bias the estimated prices for the concerned cities upwards. Before constructing a ranking of cities by their price level, I will try to mitigate that problem by excluding properties that are only on the market for a short time to capitalize on exceptionally high prices.

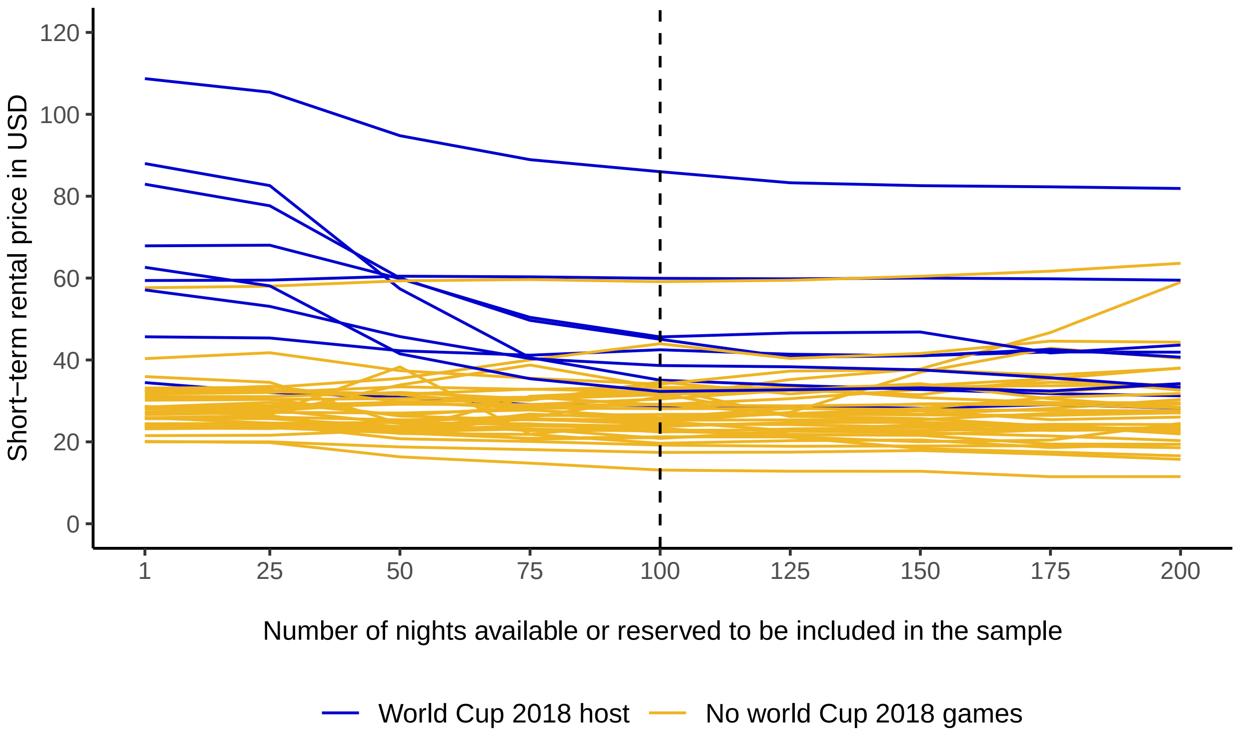
I do this by imposing a minimum number of nights in which a property is either reserved, or free and available for reservation. However, it is not evident how to choose a suitable cutoff. I use the Men's FIFA World Cup that took place in Russia from June 14th to July 15th, 2018, as a natural experiment. My sample contains 44 Russian cities, nine of which hosted games during the tournament.[21]

I run the first-stage regression on multiple subsets with increasingly strict cutoffs for the minimum number of nights on the market. The first set includes all properties in my sample of cities. The second subset only includes properties reserved or available for at least 25 out of 365 nights. I then proceed in steps of 25 nights, eventually reaching the strict requirement of 200 nights. For each regression and all Russian cities in the sample, I estimate the price for a representative property in the city center as defined above.

Figure 2 shows the results of this exercise. It depicts Russian cities that hosted World Cup games in blue and cities that did not host world cup games in yellow. Host cities are inherently different. Prices in these cities are higher even for properties on the market for most nights. This regularity makes intuitive sense, as games are usually played in larger cities that can provide the required infrastructure. However, more relevantly, the price gap between host and non-host cities decreases in the number of nights on the market. It is most prominent for the whole sample without restrictions and then declines monotonically for most pairs of cities. Depending on the city, removing properties that were only on the market for less than 25, 50, or 75 nights leads to a substantive drop in the estimated price. The decline then fades out, with a modest change associated with removing properties that were reserved or available between 75 and 100 nights. Removing properties that were on the market beyond 100 nights does not change the estimated prices in any significant way, with the blue lines becoming essentially horizontal. With at least 100 nights on the market, the remaining properties will hardly be inhabited by ordinary long-term tenants capitalizing on major

[21] In total, the world cup was played in 12 stadiums, with Moscow featuring two venues. However, Saransk and Sochi do not meet the cutoff of 300,000 inhabitants to be included in the sample.

Figure 2: Price of a representative apartment at the city center: Russia



Note: The figure shows nightly US Dollar prices of a representative short-term rental property at the city center for the 44 Russian cities in my sample. I estimate the prices by taking the exponential of city-fixed effects that are the output of 9 different hedonic regressions. These hedonic regressions differ in the subset of Airbnb properties they consider. The 44 coefficients at the right end of the x-axis are based only on properties that were available or rented on at least 200 of the last 365 nights before a property was last scraped. When moving further left on the x-axis, weaker cut-off values apply. Cities that hosted games during the Men's Fifa World Cup 2018 are shown in blue, while cities that did not are depicted in yellow.

price surges.

Non-host cities show no general trend along the whole spectrum of subsets, with the estimates becoming slightly more dispersed as the considered properties get scarcer. While these cities certainly also have varying demands over the year, only major events seem to lead to a notable rise in Airbnb properties supplied that can explain the pattern of prices. It is reassuring that one non-host city also features higher prices over the year while still displaying a stable estimated price across the different subsets. This city is Vladivostok, which was disregarded as a venue to have shorter travel distances (FIFA, 2010).

As a consequence of this analysis, I limit the data to properties that were reserved or available at least 100 out of 365 days. Due to this restriction, I lose about half of the properties, leaving me with slightly more than 1.53 million observations.

3.2 First-stage results

I then continue to estimate equation (1). Table A1 lists all 734 cities by their nightly short-term rental rate of a representative property at the city center. As defined above, that property can be rented in its entirety by a maximum number of two guests, has one bedroom and one bathroom, and corresponds to the average within its city regarding all other characteristics. Given the data and methodology I use, the most expensive city is Amsterdam, with a nightly rate of 252 USD, followed by San Francisco (243 USD), London (231 USD), and New York (225 USD). My second-stage regressions are based on (the logs of) these prices.

While these cities on top of the list are all infamous for high housing prices, they are also major tourist destinations. Perhaps as a consequence of this combination, all four cities introduced some regulation regarding Airbnb properties early on (see, for example, von Briel and Dolnicar, 2021). I cannot rule out that differences in the strictness of these regulations influence prices for such illustrious cities. However, given this study's large number of cities, I do not expect this to be a significant issue for my second-stage analysis. To control for the exposure to tourism, I include the number of Airbnbs per 1,000 inhabitants as a control variable in the second stage.

The right column of Figure A1 shows the estimated coefficients of the control variables for the first-stage hedonic regression. The number of bedrooms, bathrooms, and maximum allowed guests all increase the price of a property monotonically, although the differences between the individual coefficients vary. For example, a host can charge substantially more if she allows two guests instead of one. Hosting three instead of two guests increases prices by much less. Compared to the other coefficients, the type of the listing has a large effect on prices, with shared rooms coming with a particular markdown.

More photos in the ad correlate with higher prices, with a slightly diminishing marginal return. Apartments offered by hosts with several properties on the platform are more expensive. These hosts might learn to optimize traveler experience when spending a lot of time on the platform, allowing them to charge higher prices. They might also be able to charge more if they have market power on particular submarkets. Finally, properties close to an ocean or big lake are more expensive.

Figure A2 assesses the effects of various amenities. The amenities most positively related to price are air conditioners, pools, and TVs. There are also amenities that are negatively correlated with price. The heterogeneity of these cases suggests there are several different explanations for this. Examples include situations where the necessity of an amenity points towards an inconvenience, like a lock on the bedroom door or room darkening shades, as well as amenities that signal that an apartment is not optimized for travelers, like children's books and toys or a washer.

3.3 Robustness checks

I perform extensive checks to assess the robustness of my first-stage results. For each version, I recalculate a ranking with the estimated prices at the city center. Table A2 reports the correlations between the prices estimated with the different specifications, while Table A3 shows corresponding Spearman's rank correlations.

Specification A denotes the baseline version described above. Specification B restricts the sample to the category "entire home/apartment," dropping listings classified as "private room" (29% of all listings) and "shared room" (2%). Entire apartments are certainly what first comes to mind when thinking about the long-term rental market and might therefore appear to provide the closest correspondence. However, people also share apartments or rooms for extended periods, a prominent example being university students.

The data on Airbnbs also contain information about customer ratings, albeit for a reduced number of properties (74% of all properties in my sample of cities, but 90% of properties that were reserved or available at least 100 out of 365 days). Specification C controls for the demeaned ratings in the following categories: accuracy, check-in, cleanliness, communication, and value. I abstain from including the location rating as it might interfere with the distance gradients. Specification D controls for all amenities in the data, even if they are available only in very few properties.

There is a slight subtlety concerning the control for a location within 500m of an ocean or big lake: I demean all variables other than the number of bedrooms, bathrooms, the maximal number of guests, and the listing type by city. This implies that the representative property located at the city center has the characteristics of an average property in the city. However, when it comes to proximity to a large water body, the city center is either located close to a big lake or ocean, or it is not. In that sense, controlling for demeaned water proximity implies a somewhat flawed interpretation. On the other hand, omitting it means ignoring a factor that considerably impacts real estate prices while being correlated with proximity to the city center. Therefore I report both, with specification E omitting demeaned water distance.

Specifications F and G revisit the exclusion of properties that have been on the market for less than 100 of 365 days. Specification F includes all properties reserved at least once and thus having a revealed price. Including them approximately doubles the number of properties I can use to compute the first-stage regressions. However, it can lead to overestimating prices for cities with important events during a limited number of days. In contrast, specification G applies a more stringent requirement of 125 days on the market, resulting in a drop of another 240,000 properties. A few prices seem unrealistically high and are most likely erroneous. Therefore, my main specification winsorizes prices to each country's 0.01 and 0.99 percentiles. Specification H uses the prices at face value without winsorizing.

Specifications I and J refer to the choice of the city center. As described in Section 2 and in Nöbauer (2023), my city centers are mainly based on city tags set on OpenStreetMap using crowd intelligence. Together with a research assistant I evaluated all of these centers. Whenever they do not withstand a visual assessment, I continue with Google Maps city coordinates. If these are also suboptimal, I propose my own best guess. In contrast, specification I uses the coordinates from OpenStreetMap for all cities, while specification J uses the coordinates from Google Maps. Moreover, specification K uses ln(distance) instead of ln(distance + 1) to compute the distance gradients.

Specification L introduces a new way of demeaning. It demeans the same variables as the baseline specification. However, instead of demeaning them by city, I split the properties in each city into two subsets according to their air-line distance to the city center. I then demean the variables by city halves. This procedure is less prone to confound the effects of amenities with the distance gradients. For example, the amenity street parking negatively affects a property's price. This amenities suggests the absence of a garage or another secured parking facility. However, very central properties might not have any parking possibility and may yet be highly attractive. Therefore, part of the negative effect of street parking might be because it is a proxy for non-central locations. This problem is alleviated by demeaning within the groups of more central and less central properties in each city.

Overall, the results of the robustness tests are reassuring. The median price correlation across specifications A to L is 0.98, and the median rank correlation is 0.95.

Finally, specification M computes a ranking based on estimated average rental prices in the city, instead of estimated rental prices at the city center. It is based on the same first-stage hedonic regression as the other specifications, but it does not include distance gradients. This specification exhibits substantially lower correlations with the other specifications. However, the correlations are still 0.92 (prices) and 0.89 (rents), which might not be surprising given the large international differences in price levels.

3.4 Comparison with longterm rental data

In principle, people offering Airbnb properties compete for the same apartments as long-term renters. Living space is the primary input for the service offered by Airbnb. There is no apparent reason why other inputs like furniture or labor conducted by cleaners should vary differently between locations for the two markets. Therefore, in a market economy, we can expect Airbnb prices to be high in places with high long-term rentals and vice versa.[22] However, long-term rentals are substantially regulated in some countries, especially concerning existing tenants. Therefore, the results in this paper reflect the market for new long-term rentals more closely, as they are usually less regulated.

As a further validity test, I compare my estimates to center prices estimated using long-term rental data. I do this for France and the United States; a choice driven by data availability. Unfortunately, I do not have access to property-level long-term rental data, so I rely on aggregated data on a granular geographic dimension.

For France, I work with *la carte des loyers*.[23] This map is provided by the French government

[22] The period covered by my data ensures that the estimates are not influenced by the Covid pandemic with all its implications, which were very different for the two sectors.
[23] The data can be found on `https://www.data.gouv.fr/fr/datasets/carte-des-loyers-indicateurs-de-loyers-`

and is based on 7 million real estate ads posted between 2018 and 2022 on `seloger.com` and `leboncoin.fr`. It displays prices estimated by a hedonic regression for 34,980 French communes and arrondissements.[24] 451 of these geographical units are located in one of the 12 (functional) French cities in my study.[25] I regress the log of the rents from *la carte des loyers* on a city intercept and a city distance gradient for each city in my sample. Panel A of Figure A3 presents the results. The axes depict the city intercepts, which can be interpreted as the log prices for representative properties at the city center, once estimated using data from Airbnbs and once from *la carte des loyers*. The correlation is relatively high. Both sources also consistently estimate the extent to which Paris is an outlier in the French context; a regularity that is also found by Combes et al. (2018).

For the United States, I use data from the American Community Survey. The data cover 2015-2019 and are spatially disaggregated at the block group level.[26] I regress the log of median rents on a city intercept and city gradient for each city. Unlike above, the rents are summary statistics from survey responses rather than the result of a hedonic regression. Therefore, I control for a list of covariates linked to real estate at the block group level.[27] Similar to above, Panel B of Figure A3 compares the city intercepts estimated using long-term and short-term rental data for the United States. There is again a clear positive correlation between the two, albeit it is less clear-cut than that for France. One explanation for this discrepancy could be that I have data on all types of renters in the US, not only for apartments currently on the market. Taking long-term tenants into consideration implies a larger impact of rent control or subsidized housing, with differences in the extent of such programs and rules between cities. Moreover, the US has more local autonomy regarding taxes and public services than France. More remote places might be attractive for institutional reasons, which can impact gradients and, indirectly, the estimated prices at the city center. However, overall, the mapping between prices estimated using Airbnb properties on the one hand and long-term rental data on the other seems reasonably good.[28]

`dannonce-par-commune-en-2022/` (downloaded on 2023-02-17.)

[24] This constitutes the complete universe of French communes except for 17 communes in Mayotte. For the large cities of Paris, Marseille, and Lyon, the information is available at the level of arrondissements (neighborhoods). Their hedonic regression accounts for surface area, average surface per room, as well as year, trimester, and source of the ad.

[25] I spatially join the centroids of the communes and arrondissements to the city polygons. I drop an additional six communes, which are within the extent of my cities but do not host any Airbnb that was on the market for at least 100 days.

[26] The data cover 219,773 block groups. 80,550 of these block groups (measured at their centroid) are within one of the 70 US cities in my sample. I further restrict the analysis to the 43,636 block groups that host an Airbnb, which meets the minimum criterium of 100 nights available or reserved.

[27] For each block group, I have information about the fraction of apartments that meet certain brackets in the following categories: bedrooms, units in the building, construction year of the building, and year the tenant moved in. Moreover, I control for the fraction of apartments with a kitchen, with plumbing, and for whether the block group borders an ocean or a big lake.

[28] In the case of the US, there are three outliers in Sandy, West Valley City, and Overland Park. These are three of the very few cases in which the global rules, according to which I delimitate cities, lead to suboptimal outcomes. Overland Park might be more accurately described as part of Kansas City, while Sandy and West Valley City should probably form a single city with Salt Lake City.

4 Second-stage regressions

In the second stage, I regress the logs of the prices obtained in the first stage on log city size while controlling for a list of city characteristics. The literature on agglomeration effects typically uses either population or population density to explain wage differentials across cities. Henderson et al. (2021) test more elaborate density measures, but find that they do not offer a real improvement over simply using population density. I mostly follow Combes et al. (2018) in using the log of population size as my primary variable of interest while controlling for log area. This approach can be seen as an unrestricted version of population density, in which the coefficients of population and area are not coerced to be the opposite of each other. Combes et al. (2018) also provide an economic intuition to this approach: Controlling for area is the equivalent of restricting a city from expanding outwards when it is confronted with a higher population size. Correspondingly, they find city size to increase real estate prices more strongly when they control for area, compared to when they do not.

Table 1: Summary statistics, variables of second-stage regressions

Variable	Mean	SD	Q10	Median	Q90	N
Price of representative apartment	54.10	35.98	23.11	40.50	105.49	733
Population in 2015	2,102,975	3,690,911	359,257	906,728	4,219,852	733
Population in 1975	1,018,667	1,854,413	141,651	473,417	2,029,259	733
Area in km2	404	603	83.81	215	846	733
Compactness	0.72	0.11	0.57	0.74	0.85	733
Elevation in m	327	557	13.40	79.68	1,126	733
Difference to 21.11°C	6.74	4.45	1.39	5.79	13.01	733
Located by ocean or big lake	0.35	0.48	0.00	0.00	1.00	733
Capital	0.16	0.36	0.00	0.00	1.00	733
Airbnbs per 1,000 inhabitants	2.77	4.48	0.16	1.04	7.77	733
In 50 cities with most homicides	0.06	0.24	0.00	0.00	0.00	733
Homicides per 100k (Mex)	36.57	33.22	6.10	27.51	86.43	38
Borders USA (Mex)	0.08	0.27	0.00	0.00	0.00	38

Table 1 presents summary statistics of the variables used for these second-stage regressions. Elevation and temperature also come from the Global Human Settlement project.[29] For temperature, I follow Chauvin et al. (2017) in considering the difference to 21.11°C, which they characterize as

[29] They are included in a dataset that describes their urban centers. In some instances, I split these urban centers into more than one city (see Section 2 and Nöbauer, 2023). While I can precisely compute the area and estimated population size for these divided cities, I have the data on temperature and elevation only for the entire urban centers. In the case of split cities, I assign the values of the underlying urban center to all cities that emerge through these splits.

the “middle ground within the [...] range that is often discussed an ideal for human comfort” (p. 27).[30] I lose one observation (Weihai in China) because of a missing value regarding temperature. Capital is an indicator variable for whether a city is the national capital of its respective country, while “[l]ocated by ocean or big lake” is an indicator for whether a city borders a major water body. The first-stage hedonic regression already includes an indicator variable for whether a property is located close to an ocean or big lake ($\geq 80\text{km}^2$). However, besides influencing the price of individual properties, being located at a shore might also impact how cities as a whole are organized and experienced. I also include a measure of the number of Airbnb properties per 1,000 inhabitants to control for the attractiveness of a city to tourists.

Moreover, using my city polygons, I compute a compactness measure based on Angel et al. (2020). It assesses how much the shape of a city resembles a circle on a scale from 0 to 1. Technically, I compute a circle with the same area as the city itself around each city’s centroid and then measure the proportion of the circle that intersects with the shape of the city (the “exchange” measure in Angel et al., 2020). The rationale for this is that accessibility is dependent not only on the size of the area in which a given population is distributed, but also on the form that area takes. A circular area makes it easier to provide a high level of accessibility from many locations than a drawn-out or ramified one. The differences in accessibility can, in turn, affect how much people are willing to pay to live in the city center. Figure A4 shows the measure for four exemplary cities corresponding to the highest compactness value, the 75% quantile, the 25% quantile, and the lowest compactness value in the sample.

4.1 Second-stage results

Table 2 shows the main results of my second-stage regressions. All six specifications include country-fixed effects, which implies that I estimate the elasticity of housing costs with respect to city size from within-country variation. The population coefficient is statistically significant at the 1% level in all OLS specifications. Without any controls, I estimate an elasticity of housing costs with respect to city size of 0.139. Once I control for area, this coefficient increases to 0.164. This implies that the association between population size and housing costs is stronger when cities are not allowed to expand outwards. In that case, every additional person must be absorbed by infill (less green space or vacant plots within the city), vertical growth (taller buildings), or reduced living space per person (smaller housing units or more people per housing unit). The difference between the population coefficients under the two settings is 0.025 without additional controls, but it increases to 0.042 once the other controls are introduced.

Column 4 shows my preferred specification. It reports an elasticity of housing costs with respect to city size of 0.161. In other words, if the population size of a city increases by 10%, housing costs rise by 1.61%. This global estimate is somewhat smaller than the estimates of the elasticity of house prices with respect to city size that Combes et al. (2018) report for France. Their estimates

[30] They separately consider temperature differences in January and July, while I only have annual averages.

Table 2: Main specifications

Dependent Variable:	log(Price of representative apartment at city center)					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	IV	IV
log(Population)	0.139***	0.164***	0.119***	0.161***	0.118***	0.153*
	(0.015)	(0.057)	(0.013)	(0.050)	(0.016)	(0.084)
log(Area)		-0.030		-0.050		-0.042
		(0.063)		(0.053)		(0.087)
Compactness			-0.045	-0.054	-0.045	-0.053
			(0.118)	(0.112)	(0.118)	(0.113)
Elevation (100m)			-0.007*	-0.007*	-0.007*	-0.007*
			(0.004)	(0.004)	(0.004)	(0.004)
Difference to 21.11°C			0.002	0.002	0.002	0.002
			(0.007)	(0.007)	(0.007)	(0.007)
By ocean / big lake			0.055	0.054	0.055	0.055
			(0.043)	(0.045)	(0.043)	(0.045)
Capital			0.106*	0.105*	0.108*	0.106*
			(0.057)	(0.057)	(0.058)	(0.059)
Airbnbs per 1,000			0.030***	0.029***	0.030***	0.029***
			(0.006)	(0.006)	(0.006)	(0.006)
Cragg-Donald F-Stat					2,959.6	268.0

Note: The table shows regressions of the estimated price of a representative short-term rental property at the city center on city size and control variables. The units of observation are 733 cities. All specifications include country fixed effects. Population in 2015 is instrumented by population in 1975 for the IV specifications. The parentheses show standard errors, which are clustered by country. The levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

range from 0.176 to 0.305, with their preferred estimate being 0.208. At the same time, their estimated area-unrestricted elasticity is 0.109, which is almost precisely what I find. The fact that the difference manifests itself in the area-restricted elasticity could be consistent with French cities being more limited in vertical growth by stricter regulations than cities elsewhere. However, it is important to keep in mind that my sample consists of cities that are, on average, more than 12 times larger than the cities used by Combes et al. (2018). They also estimate an elasticity that is non-linear in population size and find (area-restricted) estimates as large as 0.288 of a city with one million inhabitants and 0.378 for a city as large as Paris. Comparing these estimates with mine suggests that the housing costs in big French cities increase faster in city size than the housing costs in big cities elsewhere.

Columns 5 and 6 mirror the specifications of columns 3 and 4, but introduce an instrumental variable approach. I instrument log population in 2015 with log population in 1975. The idea is that today's price level (and other recent developments that affect it) might cause people to move into (or away from) a given city and, might therefore, bias the estimation. At the same time, long-past population counts should be unaffected by it. In applying this strategy, I follow a standard approach introduced by Ciccone and Hall (1996) and used amongst others by Combes et al. (2008) and Combes et al. (2018). Unlike these papers, my work deals with a worldwide sample, and, unfortunately, it is impossible to find ancient population counts on that scale. The advantage of the 1975 population data I use, apart from its existence, is that it is provided by the same source, built using the same principles, and covering the same grid as the 2015 population data. However, 40 years are not enough to alleviate concerns about the instrument's validity. As Chauvin et al. (2017), who use population data from 1980 to construct an IV, I argue that columns 5 and 6 should not be interpreted as more than a robustness check. The point estimates are almost unchanged between columns 3 and 5, with the variable of interest still being statistically significant at the 1% level. When I control for area, the coefficient of log population decreases from 0.161 to 0.153 between the OLS and the IV estimation. It is only statistically significant at the 10% level in the IV setting, compared to the 1% level with OLS.

Concerning the control variables, I estimate capital cities to be about 10.6% more expensive than other cities, with the effect being statistically significant at the 10% level. A higher number of Airbnbs is associated with higher prices. This relation is statistically significant at the 1% level. The predicted housing cost difference between a city at the 25% quantile and a city at the 75% quantile of Airbnb properties is 0.079. A statistically significant (at the 10% level) relation exists between elevation and housing costs. However, given that the average elevation is 328 meters, with a median of 80 meters, this effect is quantitatively small. Moreover, I estimate more compact cities to be cheaper and cities at the seaside to be more expensive beyond the properties close to the shore. However, neither of these effects is statistically significant.

Importance of fixed effects

Table A4 explores the explanatory power of the different sets of variables. The first column regresses the log price of the representative apartment at the city center merely on the logs of population and area. Without country-fixed effects, the direction of the effect switches. This behavior is consistent with the fact that lower income countries tend to have denser cities with less living space per person (Jedwab et al., 2021). Specification (2) consists of the control variables only. The point estimates of the controls go in the same direction as in the full specification, but they are larger, which can be explained by the omission of the country-fixed effects. Log population and log area alone (column 1) and the controls alone (column 2) explain an R^2 of around 0.3.

The R^2 increases to 0.48 in column (3), which includes both the logs of population and area and the controls. Adding controls without country-fixed effects still results in a negative correlation

Table 3: Heterogeneity by country

Dependent Variable:	log(Price of representative apartment at city center)						
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	USA	Eurozone	Russia	China	India	Brazil	Mexico
log(Population)	0.23**	0.33***	0.19	0.11	0.43**	0.05	-0.36**
	(0.10)	(0.08)	(0.18)	(0.12)	(0.16)	(0.09)	(0.17)
log(Area)	-0.07	-0.13	0.09	0.02	-0.25	0.05	0.52**
	(0.12)	(0.10)	(0.22)	(0.13)	(0.19)	(0.11)	(0.19)
Compactness	-0.38**	0.33	-0.30	-0.32	-0.22	0.35	-0.74
	(0.18)	(0.32)	(0.43)	(0.32)	(1.21)	(0.23)	(0.50)
Elevation (100m)	-0.03***	-0.03	-0.03	0.00	0.00	0.00	0.00
	(0.01)	(0.03)	(0.03)	(0.01)	(0.03)	(0.01)	(0.01)
Difference to 21.11°C	0.02***	0.08***	0.00	-0.01*	-0.05	-0.01	-0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.05)	(0.01)	(0.02)
By ocean / big lake	-0.12**	-0.03	0.04	0.20**	0.19	0.03	0.08
	(0.05)	(0.06)	(0.07)	(0.08)	(0.27)	(0.07)	(0.14)
Airbnbs per 1,000	0.04***	0.02***	0.05**	0.04***	0.74**	0.10***	0.10***
	(0.01)	(0.00)	(0.02)	(0.01)	(0.34)	(0.03)	(0.02)
Country fixed effects	-	Yes	-	-	-	-	-
Observations	70	76	44	112	31	44	38

Note: The table shows regressions of the estimated price of a representative short-term rental property at the city center on city size and control variables. The units of observation are cities. The parentheses show standard errors clustered by country for specification 2 (eurozone) and heteroscedasticiy robust standard errors for all other specifications. The levels of significance are * p < 0.10, ** p < 0.05, *** p < 0.01.

between population count and price.[31] Specification (4) exclusively contains country-fixed effects and shows that they alone generate a R^2 of 0.74. Columns (5) and (6) mirror specifications (2) and (4) of Table 2. Adding the logs of population and area increases the R^2 to 0.78, while additionally adding controls raises it to 0.81.

Geographic heterogeneity

Table 3 repeats the analysis for the six countries with the highest number of cities in the sample, and for the eurozone. Given the low number of observations, these results should be taken with a grain of salt. Nevertheless they show some interesting regularities.

I estimate a statistically significantly positive elasticity for the United States, India, and the euro-

[31] An obvious omitted control variable in this specification is income. I do not include it since all other second-stage regressions either include country-fixed effects or focus on one particular country. Spatially disaggregated within-country data on income per capita is not readily available on a global scale.

zone. In all cases, the estimated elasticity is higher than for the full global sample, with particularly large effects in the latter two. While the elasticity is also somewhat higher in Russia, the estimate is not statistically significant. My results for China and Brazil show positive estimates that are below the world-average and not statistically significant. Mexico is the only country that completely falls out of line. It shows a statistically significant effect that is negative and large. This implies that city size is negatively correlated with housing costs in the Mexican context. I will come back to this below.

As stated above, Combes et al. (2018) find an elasticity of 0.29 for a city with one million inhabitants in the French context. The average eurozone-city in my sample has a population of 1.07 million, which makes that a good comparison. Estimating a model with country-fixed effects for the eurozone, I find a coefficient of 0.33. It is hard to check whether France is representative for the eurozone. If it is, this finding would suggest that the elasticity of housing costs with respect to city size is indeed increasing in city size. Moreover, the estimated elasticity from Combes et al. (2018) might be on the higher end of the global spectrum.

There are at least two plausible explanations for the lower and not statistically significant estimates for Russia, China, and Brazil and for the fact that the coefficient of log area is estimated to be positive. First, these countries might have less stringent regulations concerning building heights or building over green spaces. They might also simply have more room for infill. Second, the results might also be biased towards zero because of data quality issues. While China has the highest number of cities in the sample, delineating the cities and setting their center was harder than anywhere else. For example, the maps displayed by Google Maps are not superimposable to satellite images for Chinese cities because of government regulations. Instead they are shifted in a non-monotonic way (Fuentes, 2019). In contrast, the United States and the eurozone have higher numbers of Airbnbs in the sample than all other countries mentioned in Table 3. This might lead to more precisely estimated housing costs in the first-stage.

While Table 3 does not include capital dummies, Table A5 repeats the analysis without the two largest cities for each entity.[32] Most of the results are very robust to the exclusion of these cities. The notable exception is Russia, where the coefficient drops from 0.19 to 0.02 after the exclusion of Moscow and Saint Petersburg, implying that the positive relation between city size and housing costs is entirely driven by these two metropolises.

Comparison with Chauvin et al. (2017)

Chauvin et al. (2017) also provide recent estimates of the elasticity of housing costs with respect to city size for multiple countries. They focus on other aspects of the spatial equilibrium, amongst others on agglomeration economies. However, an appendix to their paper includes such estimates for the United States, Brazil, China, and India. They estimate their regressions using OLS and IV

[32] The excluded cities are New York, Los Angeles, Paris, Madrid, Moscow, Saint Petersburg, Shanghai, Beijing, Delhi, Mumbai, São Paulo, Rio de Janeiro, Mexico City, and Guadalajara.

specifications, based on data from 2010 and with the population in 1980 as an instrument.[33] This is close enough to the years I use (population data from 2015, short-term rental prices from 2018-19, population from 1975 as an instrument) to expect similar results.

Table A6 presents this comparison. All point estimates correspond to the effect of log population. My preferred OLS specification includes the same controls as Table 3 but excludes area to be consistent with the estimates from Chauvin et al. (2017).[34] My preferred IV specification additionally instruments log population in 2015 with log population in 1975. Our results are very similar for the US, where the data availability is the best.[35] Chauvin et al. (2017) present a specification with log rent and another one with log price as the dependent variable, and my estimates fall right in between the two. I get somewhat lower point estimates for Brazil while confirming the positive and statistically significant elasticity of housing costs with respect to city size. They also report two separate regressions for China. My preferred estimates are again between the two estimates of Chauvin et al. (2017).

The similarity of the results disappears in the case of India. Chauvin et al. (2017) find no statistically significant effect of city size on housing rents for India, with point estimates narrowly above and below zero. They do find agglomeration economies for India that are about 50% higher than for the US, which implies that real wages must increase in city size. They explain this with low migration rates and geographical differences in the level of education but also acknowledge that the data quality of their rent data might offer another explanation (p. 32). In contrast, I do find a statistically significant positive relation between short-term rental prices and city size. The corresponding coefficient is about 35% larger than that for the United States. If amenities increase less (or decrease more) with city size than in the US, this could very well be in line with the standard spatial equilibrium model whose applicability to the Indian context is challenged by Chauvin et al. (2017).

There are some notable methodological differences concerning the estimation of this elasticity. First, Chauvin et al. (2017) do not report to account for property-level characteristics, while my rental price indices are the outcome of hedonic regressions. Second, I estimate prices at the city center, while Chauvin et al. (2017) appear to use city fixed effects without accounting for any geographical within-city dimension. They also do not use city-level controls in the second stage. If I adjust my methodology concerning the second-stage regression, I also estimate elasticities for India that are very small and statistically indistinguishable from zero. While the same pattern emerges for the US and China, albeit to a smaller degree, the results for Brazil are unchanged (strips 3 and 4 of Table

[33] They also include IV estimates based on older population counts: 1900 for the United States, 1920 for Brazil, 1950 for China, and 1951 for India. I restrict my comparison to their first set of IV estimates as it provides better comparability.

[34] They also report coefficients of regressions using density as the independent variable. Those results are qualitatively similar, except for house prices in China, where they report statistically significant effects of around 0.22.

[35] Chauvin et al. (2017) work with household level data and report the following sample sizes: 24.4 mio / 44 mio (rent/price) for the US, 818 k for Brazil, 6.7 k / 25 k (rent/price) for China, and 3.3 k for India. My work builds on the following numbers of Airbnb properties that were available at least 100 out of 365 nights: 282 k for the US, 36 k for Brazil, 169 k for China, and 9.4 k for India.

A6).

Homicides and the elasticity of housing costs with respect to city size

Why are the results that different for Mexico? One possible explanation could be the level of crime. Glaeser and Sacerdote (1999) argue that "it is ironic that the same urban advantages, lower transport costs, faster urban information flows, and the same scale economies that help to make cities more productive also increase the level of crime in the city" (p. 241). Safety considerations might not affect the attractiveness of cities too much when the overall level of crime is low. However, Mexico is in a drug war and experienced over 72,000 homicides in 2018 and 2019 alone.[36] It seems plausible that population density can seem frightening in such an environment. In a Roback (1982) type setting, crime can act as a negative amenity, with crime-ridden places having to offer higher wages or lower real estate prices in equilibrium. If the probability of becoming the victim of a crime increases in city size, this can explain why the positive relationship between city size and real estate prices might not hold in places with high crime rates.

Table 4 explores this dimension. Column (1) reports the baseline regression for the 38 Mexican cities in my sample.[37] I include a dummy for whether a city borders the United States since I expect the Mexican real estate market and potentially also crime rates to be affected by proximity to the US. Column (2) includes the homicide rate per 100,000 inhabitants and an interaction term between the homicide rate and population size. The data originate from the "Instituto Nacional de Estadística y Geografía" and are cleaned and made available by Diego Valle-Jones.[38] I use the average of the 2018 and 2019 homicide rates. The yearly average homicide rate among the 38 Mexican cities in my sample is 36.6 per 100,000 inhabitants (see Table A4), while the average population size of these cities is 1.46 million.

I find a negative interaction between homicides and population size that is statistically significant at the 10% level. This result implies that the negative correlation between short-term rental prices and city size that I find in the Mexican context is particularly strong for cities with a high homicide rate. While controlling for the homicide rate makes the baseline effect of population size smaller and statistically insignificant, its point estimate is still negative. However, if crime is indeed a driver of this reverse effect, it is plausible that even the safer cities in Mexico are affected to some degree.

I, therefore, try to go beyond Mexico. Column 3 of Table 4 replicates the baseline specification for the entire worldwide sample. Column 4 adds an indicator for whether a city appears in the 2018 or 2019 versions of the list of the 50 cities with the highest homicide rates that is published each year by the "Consejo Ciudadano para la Seguridad Pública y la Justicia Penal AC".[39] The baseline

[36] This number is based on the same data from the INEGI that I describe below.
[37] The crime data are based on metro areas and includes several big municipios that are not part of a metro area. For details, see `https://github.com/diegovalle/mxmortalitydb`, last accessed on 2023-06-11. I match these metro areas to my cities by name and verify that they include the same center. However, the boundaries of the metro areas and my cities differ to some extent.
[38] I last downloaded the data on 2023-06-09 from `https://github.com/diegovalle/mxmortalitydb`.
[39] I downloaded the rankings on 2023-05-27 from `https://geoenlace.net/seguridadjusticiaypaz/webpage/`

Table 4: Interaction with the homicide rate

Dependent Variable:	log(Price of representative apartment at city center)			
Extent:	Mexico	Mexico	World	World
Model:	(1)	(2)	(3)	(4)
log(Population)	-0.284	-0.201	0.161***	0.164***
	(0.191)	(0.189)	(0.050)	(0.050)
log(Area)	0.419*	0.451**	-0.050	-0.050
	(0.215)	(0.218)	(0.053)	(0.054)
Compactness	-0.820	-0.919	-0.054	-0.050
	(0.491)	(0.581)	(0.112)	(0.113)
Elevation (100m)	0.002	0.004	-0.007*	-0.007*
	(0.006)	(0.006)	(0.004)	(0.004)
Difference to 21.11°C	-0.007	-0.012	0.002	0.002
	(0.020)	(0.019)	(0.007)	(0.007)
By ocean / big lake	0.058	0.089	0.054	0.054
	(0.128)	(0.145)	(0.045)	(0.044)
Capital			0.105*	0.100*
			(0.057)	(0.058)
Airbnbs per 1,000	0.106***	0.106***	0.029***	0.029***
	(0.022)	(0.024)	(0.006)	(0.006)
Borders USA	0.254*	0.382***		
	(0.149)	(0.118)		
Homicides per 100k		0.051*		
		(0.025)		
log(Population) × Homicides per 100k		-0.004*		
		(0.002)		
In 50 most homicides				0.961***
				(0.339)
log(Population) × In 50 most homicides				-0.069***
				(0.024)
Country fixed effects	-	-	Yes	Yes
Observations	38	38	733	733

Note: The table shows regressions of the estimated price of a representative short-term rental property at the city center on city size and control variables. The units of observation are cities. The parentheses show heteroscedasticity robust standard errors in models (1) and (2) and standard errors clustered by country in models (3) and (4). The levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

effect of being in this ranking is positive, which reflects the fact that most of these cities are located in upper-middle-income countries.[40] However, there is a negative interaction between city size and being one of the cities with the highest homicide rates. The corresponding regression coefficient is statistically significant at the 1% level. I estimate the elasticity of housing costs with respect to city size to decrease from 0.164 to 0.095 for the most dangerous cities according to this definition.

5 Conclusion

In this paper, I estimate the elasticity of housing costs with respect to city size. I conduct the analysis on a worldwide scale, using 733 cities with at least 300,000 inhabitants and 100 Airbnb properties. I am able to work on this international scale because I use novel data on short-term rental properties from Airbnb as a proxy for housing costs. In a first-stage hedonic regression, I estimate the price of a representative property at the center of each city. I then use these prices in a second-stage regression, regressing them on population, area, city-level controls, and country-fixed effects.

My preferred estimate of the elasticity of housing costs with respect to city size is 0.16. This is somewhat less than the estimate of 0.21 that Combes et al. (2018) find for a sample of French cities that are more than 12 times smaller on average. When not controlling for area, both our samples yield an estimate of 0.11. I find the elasticity to differ substantially by country/region, estimating a coefficient of 0.33 for the eurozone. This is in line with non-linear estimates that Combes et al. (2018) provide for a hypothetical French city with one million inhabitants (0.29) and for Paris (0.38). Assuming that French cities and other cities in the eurozone are alike, this supports their finding of a non-linear elasticity of housing costs with respect to city size.

It also suggests that large eurozone cities face above-average elasticities of housing costs with respect to city size compared to other large cities worldwide, as do cities in India. In particular, I estimate elasticities that are positive but considerably smaller and not statistically significant for Russia (in particular without Moscow and Saint Petersburg), China, and Brazil. Especially given the small sample sizes of these country regressions, I cannot rule out that data issues drive part of this discrepancy. However, given that I control for city area, I hypothesize that stricter building height regulations in the eurozone and India might play some role, by limiting how much the housing stock can adjust as a reaction to population growth. Infill development within the existing boundaries of a city can play a similar role of adjustment, where cities in the eurozone and India might contain fewer empty plots.

An alternative and perhaps complementary explanation is based on the Rosen-Roback model (Rosen, 1979; Roback, 1982). The model predicts that differences in wages, housing costs, and

---

`archivos.php`.

[40] To classify countries by income, I use the World Bank definitions, available at `https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups` (last accessed: 2023-06-09).

urban amenities counterbalance each other. If a city has high wages and low housing costs, it will most likely not have good amenities. Otherwise, residents from other cities would move, driving housing costs up and wages down. Each of the three factors has an elasticity with respect to city size. Concerning the elasticity of wages with respect to city size, empirical research tends to find larger agglomeration effects in developing countries than in high-income countries (Chauvin et al., 2017; Henderson et al., 2021). Combining this evidence with my findings on urban costs has implications about the elasticity of amenities with respect to city size. If wages in the eurozone increase less in city size and housing costs increase more, the Rosen-Roback model would predict that the quality of amenities increases more strongly in city size in the eurozone (or decreases less strongly in city size) than elsewhere.

Finally, the suggestive evidence of a negative elasticity of housing costs with respect to city size for Mexico, and the negative interaction between that elasticity and homicides, both in Mexico and worldwide, open room for future research. While Glaeser and Sacerdote (1999) find crime to increase in city size for the United States, Ahlfeldt and Pietrostefani (2019) find it to decrease in density in other OECD countries. Which types of crime affect other urban costs (and perhaps also benefits) in which contexts remains an exciting open question.

6 Bibliography

Ahlfeldt, G.M. and Pietrostefani, E., 2019. The economic effects of density: A synthesis, *Journal of Urban Economics*, 111, 93–107.

Akbar, P.A., Couture, V., Duranton, G., and Storeygard, A., 2021. Mobility and congestion in urban India, Working Paper.

Alonso, W., 1964. *Location and Land Use. Toward a General Theory of Land Rent*, Harvard University Press.

Angel, S., Franco, S.A., Liu, Y., and Blei, A.M., 2020. The shape compactness of urban footprints, *Progress in Planning*, 139, 100429.

Chauvin, J.P., Glaeser, E., Ma, Y., and Tobio, K., 2017. What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States, *Journal of Urban Economics*, 98, 17–49.

Ciccone, A. and Hall, R., 1996. Productivity and the density of economic activity, *American Economic Review*, 86 (1), 54–70.

Combes, P.P., Duranton, G., and Gobillon, L., 2008. Spatial wage disparities: Sorting matters!, *Journal of Urban Economics*, 63 (2), 723–742.

Combes, P.P., Duranton, G., and Gobillon, L., 2018. The costs of agglomeration: House and land prices in French cities, *The Review of Economic Studies*, 86 (4), 1556–1589.

Combes, P.P. and Gobillon, L., 2015. The empirics of agglomeration economies, *in:* G. Duranton, J.V. Henderson, and W.C. Strange, eds., *Handbook of Regional and Urban Economics*, Elsevier, *Handbook of Regional and Urban Economics*, vol. 5, 247–348.

Duranton, G. and Puga, D., 2004. Micro-foundations of urban agglomeration economies, *in:* J.V. Henderson and J.F. Thisse, eds., *Cities and Geography*, Elsevier, *Handbook of Regional and Urban Economics*, vol. 4, 2063–2117.

Duranton, G. and Puga, D., 2015. Urban land use, *in:* G. Duranton, J.V. Henderson, and W.C. Strange, eds., *Handbook of Regional and Urban Economics 5*, Elsevier, 467–560.

FIFA, 2010. 2018 FIFA World Cup™ bid evaluation report: Russia, Report, Fédération Internationale de Football Association.

Florczyk, A., Corbane, C., Schiavina, M., Pesaresi, M., Maffenini, L., Melchiorri, M., Politis, P., Sabo, F., Freire, S., Ehrlich, D., Kemper, T., Tommasi, P., Airaghi, D., and Zanchetta, L., 2019. GHS Urban Centre Database 2015, multitemporal and multidimensional attributes, R2019A, European Commission, Joint Research Centre (JRC). Dataset.

Fuentes, E., 2019. Why GPS coordinates look wrong on maps of China, https://www.serviceobjects.com/blog/why-gps-coordinates-look-wrong-on-maps-of-china/ (last accessed: 13.01.2023).

Gallagher, L., 2017. *The Airbnb Story: How Three Ordinary Guys Disrupted an Industry, Made Billions . . . and Created Plenty of Controversy*, Harper Business.

Glaeser, E., 2011. *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*, The Penguin Press.

Glaeser, E.L. and Sacerdote, B., 1999. Why is there more crime in cities?, *Journal of Political Economy*, 107 (6), 225–258.

Henderson, J.V., 1974. The sizes and types of cities, *The American Economic Review*, 64 (4), 640–656.

Henderson, J.V., Nigmatulina, D., and Kriticos, S., 2021. Measuring urban economic density, *Journal of Urban Economics*, 125, 103188.

Henderson, V., 2002. Urban primacy, external costs, and quality of life, *Resource and Energy Economics*, 24 (1), 95–106.

Jedwab, R., Barr, J., and Brueckner, J.K., 2022. Cities without skylines: Worldwide building-height gaps and their possible determinants and implications, *Journal of Urban Economics*, 132, 103507.

Jedwab, R., Loungani, P., and Yezer, A., 2021. Comparing cities in developed and developing countries: Population, land area, building height and crowding, *Regional Science and Urban Economics*, 86, 103609.

Mills, E.S., 1967. An aggregative model of resource allocation in a metropolitan area, *American Economic Review*, 57 (2), 197–210.

Muth, R.F., 1969. *Cities and Housing*, University of Chicago Press.

Nöbauer, B., 2023. The monocentric city model worldwide: Rent, density, and transport cost gradients in 734 cities, Working Paper.

Pope, D.G. and Pope, J.C., 2012. Crime and property values: Evidence from the 1990s crime drop, *Regional Science and Urban Economics*, 42, 177–188.

Roback, J., 1982. Wages, rents, and the quality of life, *Journal of Political Economy*, 90 (6), 1257–1278.

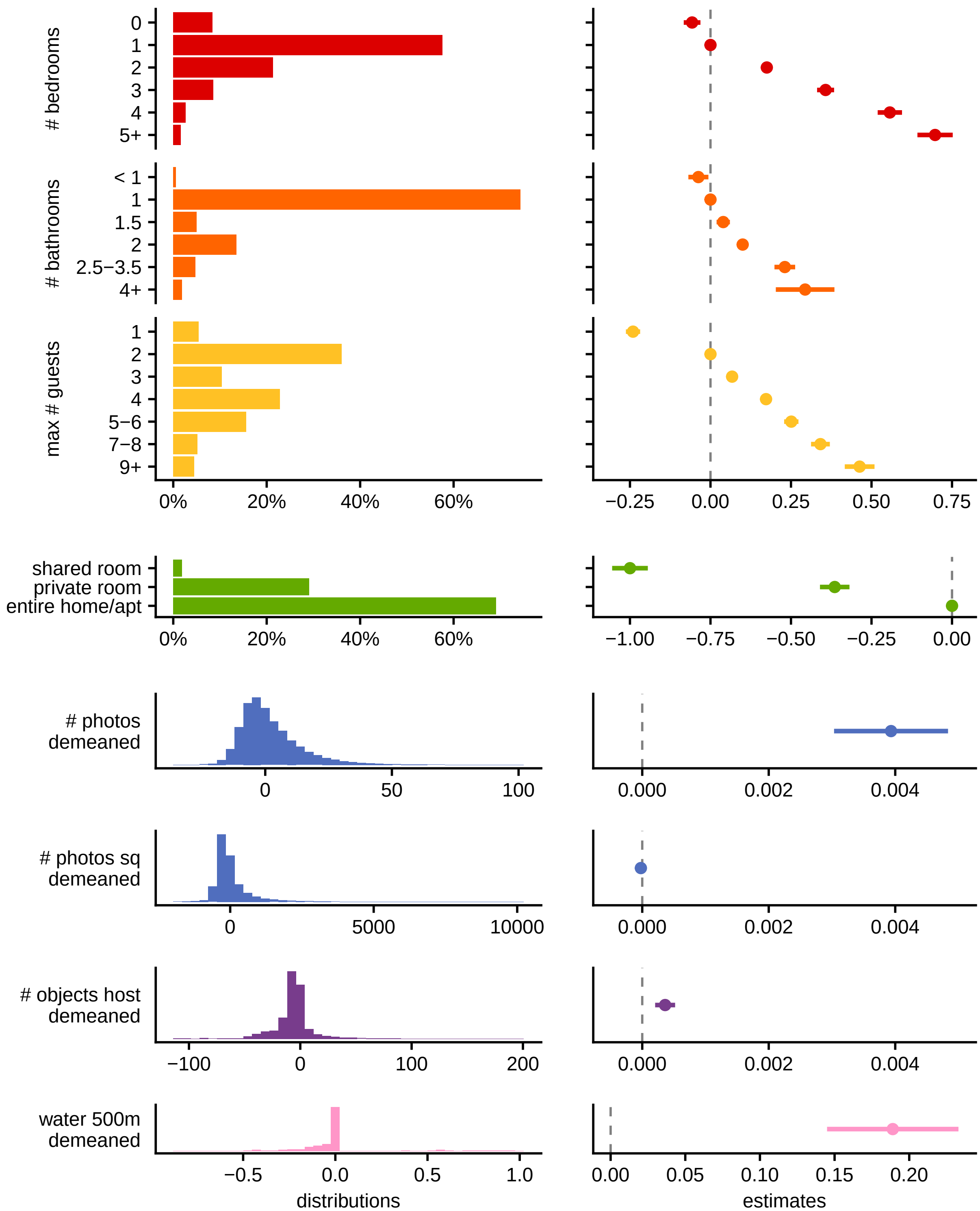
Rosen, S., 1979. Wage-based indexes of urban quality of life, *in:* P. Mieszkowski and M. Straszheim, eds., *Current Issues in Urban Economics*, John Hopkins University Press, 74–104.

Schiavina, M., Freire, S., and MacManus, K., 2019. GHS population grid multitemporal (1975, 1990, 2000, 2015) R2019A, European Commission, Joint Research Centre (JRC). Dataset.

Shirk, D. and Wallman, J., 2015. Understanding Mexico's drug violence, *Journal of Conflict Resolution*, 59 (8), 1348–1376.

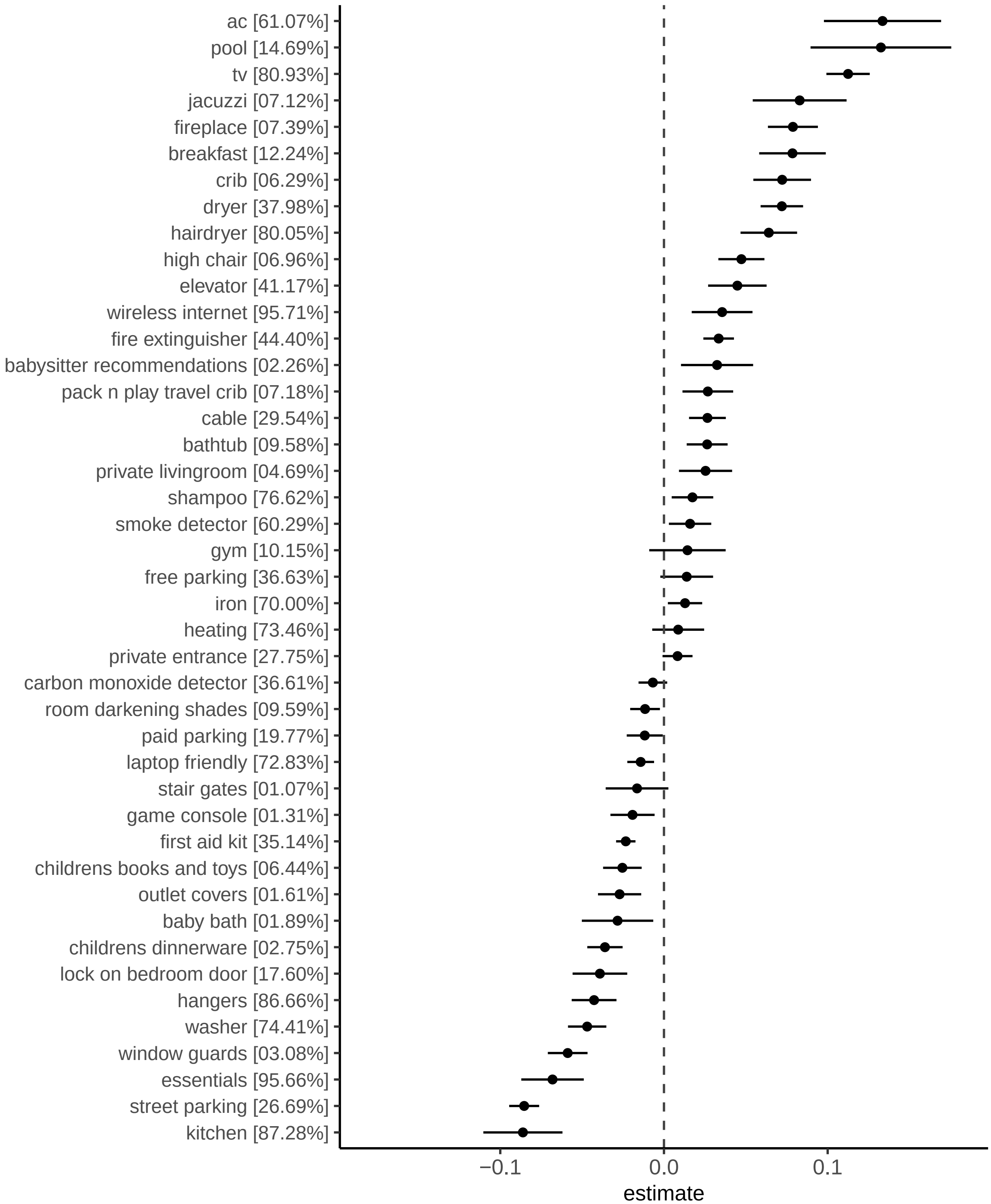
von Briel, D. and Dolnicar, S., 2021. The evolution of Airbnb regulations, *in:* S. Dolnicar, ed., *Airbnb before, during and after COVID-19*, University of Queensland.

Figure A1: Control variables of hedonic regression



Note: The left column of the figure shows the distributions of all hedonic first-stage regression control variables not classified as "amenities". For the last four panels of this column, the variables are demeaned by subtracting their respective city average. The right column shows the corresponding regression coefficients, with 99% confidence intervals based on standard errors that are clustered by city. The number of observations used in the regression is 1,532,862.

Figure A2: Control variables of hedonic regression: amenities



Note: The y-axis shows all 43 amenities which are available in at least 1% of the Airbnb properties in the sample, with the proportion of properties in which the respective amenity is available in brackets. I demean all of these amenities by city and include them as control variables in the first-stage hedonic regression. The figure shows the estimated coefficients of these variables, together with 99% confidence intervals based on standard errors that are clustered by city. The number of observations used in the regression is 1,532,862.

Table A1: Estimated nightly short-term rental rate of a representative apartment at the city center in USD

1	Amsterdam	NLD	252.26
2	San Francisco	USA	242.82
3	London	GBR	230.98
4	New York	USA	224.78
5	Austin	USA	186.78
6	Boston	USA	182.45
7	Seattle	USA	166.25
8	Washington	USA	163.31
9	Copenhagen	DNK	161.39
10	Miami	USA	160.57
11	Paris	FRA	158.82
12	Dubai	ARE	158.47
13	Kuwait City	KWT	157.47
14	Portland	USA	157.11
15	Stockholm	SWE	151.17
16	Dublin	IRL	150.95
17	Chicago	USA	149.43
18	Zürich	CHE	148.15
19	Sydney	AUS	146.44
20	Las Vegas	USA	140.62
21	Long Branch	USA	140.52
22	Roma	ITA	138.42
23	München	DEU	136.79
24	Edinburgh	GBR	136.58
25	New Orleans	USA	135.81
26	Oakland	USA	134.11
27	San Diego	USA	131.99
28	Denver	USA	131.71
29	Honolulu	USA	131.57
30	Lagos	NGA	131.31
31	Vancouver	CAN	131.17
32	Milano	ITA	130.26
33	Göteborg	SWE	130.21
34	Atlanta	USA	127.99
35	Columbus	USA	127.88
36	Oslo	NOR	125.63
37	San Juan	PRI	125.18
38	Cardiff	GBR	125.07
39	Sarasota	USA	123.59
40	Frankfurt am Main	DEU	123.48
41	Genève	CHE	123.45
42	Brighton	GBR	123.33
43	Detroit	USA	122.85
44	Liverpool	GBR	122.40
45	Basel	CHE	121.79
46	Cleveland	USA	121.79
47	Tokyo	JPN	121.41
48	Grand Rapids	USA	120.92
49	Toronto	CAN	120.58
50	Berlin	DEU	118.12
51	Indianapolis	USA	118.09
52	Louisville	USA	115.67
53	Denpasar	IDN	115.49
54	Singapore	SGP	115.28
55	San Antonio	USA	115.12
56	Charlotte	USA	115.11
57	Philadelphia	USA	114.56
58	Los Angeles	USA	114.50
59	Hong Kong	CHN	113.72
60	Pittsburgh	USA	113.19
61	Hamburg	DEU	111.48
62	Houston	USA	111.40
63	Kinshasa	COD	110.87
64	Utrecht	NLD	110.78
65	Firenze	ITA	109.86
66	Manchester	GBR	109.49
67	Barcelona	ESP	108.27
68	Kyoto	JPN	108.14
69	Fort Worth	USA	107.80
70	Jerusalem	ISR	106.49
71	Sacramento	USA	106.39
72	Milwaukee	USA	105.76
73	Palma	ESP	105.69
74	Québec	CAN	105.55
75	Tel Aviv	ISR	105.27
76	Hialeah	USA	104.85
77	Fort Lauderdale	USA	104.58
78	Cincinnati	USA	103.13
79	Minneapolis	USA	102.84
80	Bristol	GBR	102.39
81	Dallas	USA	100.34
82	Wien	AUT	100.23
83	Providence	USA	99.98
84	Kansas City	USA	99.60
85	Tucson	USA	99.30
86	Phoenix	USA	98.95
87	Orlando	USA	98.44
88	Baltimore	USA	98.31
89	Oklahoma City	USA	98.30
90	Leiden	NLD	97.63
91	Auckland	NZL	97.52
92	San Jose	USA	97.50
93	Southampton	GBR	97.41
94	Köln	DEU	97.37
95	Omaha	USA	97.22

96	Concord	USA	96.71
97	Belfast	GBR	96.14
98	Rotterdam	NLD	95.59
99	Den Haag	NLD	95.52
100	Manama	BHR	95.39
101	Jacksonville	USA	95.38
102	Des Moines	USA	95.35
103	Helsinki	FIN	95.23
104	Tampa	USA	95.07
105	Montréal	CAN	94.38
106	Bologna	ITA	94.23
107	Saint Louis	USA	94.10
108	Beijing	CHN	93.82
109	Bordeaux	FRA	93.40
110	Leeds	GBR	93.23
111	Bilbao	ESP	92.10
112	Sevilla	ESP	91.46
113	Praha	CZE	90.99
114	Bruxelles	BEL	90.94
115	Adelaide	AUS	90.38
116	Memphis	USA	90.04
117	Fresno	USA	89.50
118	Düsseldorf	DEU	89.38
119	Colorado Springs	USA	88.78
120	Melbourne	AUS	88.70
121	Glasgow	GBR	88.66
122	Antwerpen	BEL	88.55
123	Saint Petersburg	USA	88.48
124	Salt Lake City	USA	88.17
125	Nürnberg	DEU	87.71
126	Stuttgart	DEU	86.97
127	Madrid	ESP	86.65
128	Strasbourg	FRA	86.27
129	Newcastle upon Tyne	GBR	86.23
130	Birmingham	GBR	86.02
131	Moscow	RUS	85.98
132	Port of Spain	TTO	85.33
133	Dayton	USA	84.90
134	Cartagena	COL	84.87
135	Osaka	JPN	84.67
136	Lyon	FRA	83.63
137	Kolkata	IND	83.34
138	Leicester	GBR	83.20
139	Sapporo	JPN	83.03
140	Rochester	USA	82.18
141	Bakersfield	USA	82.06
142	Ottawa	CAN	81.71
143	Norfolk	USA	81.69
144	Nottingham	GBR	81.48
145	Cape Town	ZAF	81.01
146	Jeonju	KOR	80.89
147	Gold Coast	AUS	80.84
148	Nice	FRA	80.81
149	Accra	GHA	80.38
150	Buffalo	USA	80.02
151	Hannover	DEU	79.29
152	Luanda	AGO	78.85
153	Brisbane	AUS	78.22
154	Sheffield	GBR	77.85
155	Lille	FRA	77.51
156	Kingston	JAM	77.27
157	Aurora	USA	77.07
158	Bergamo	ITA	77.05
159	Xiamen City	CHN	77.01
160	Tulsa	USA	76.91
161	Tainan	TWN	76.72
162	Lisboa	PRT	76.09
163	Sendai	JPN	75.77
164	Ogden	USA	75.64
165	Málaga	ESP	75.59
166	Kanazawa	JPN	75.26
167	Augsburg	DEU	74.98
168	Calgary	CAN	74.95
169	Albuquerque	USA	74.83
170	Asahikawa	JPN	74.43
171	Bloemfontein	ZAF	74.24
172	Hull	GBR	74.04
173	Bonn	DEU	73.72
174	Tallinn	EST	73.64
175	Fukuoka	JPN	72.94
176	Marrakesh	MAR	72.63
177	Portsmouth	GBR	71.51
178	Bremen	DEU	71.40
179	Dresden	DEU	70.67
180	Overland Park	USA	70.67
181	Leipzig	DEU	70.11
182	Naha	JPN	69.65
183	Porto	PRT	69.21
184	Zhoushan	CHN	69.08
185	Winnipeg	CAN	68.93
186	Perth	AUS	68.73
187	Torino	ITA	68.65
188	Coventry	GBR	68.21
189	Ensenada	MEX	68.08
190	Beirut	LBN	67.72
191	Bangkok	THA	67.36
192	Seoul	KOR	67.34
193	Stockton	USA	67.31

194	Haiphong	VNM	67.10	243	Marseille	FRA	58.73
195	Hangzhou	CHN	66.98	244	Athens	GRC	58.58
196	Granada	ESP	66.17	245	Bayuquan	CHN	58.48
197	Napoli	ITA	65.97	246	Qingdao	CHN	58.42
198	Chaozhou	CHN	65.89	247	Abu Dhabi	ARE	57.72
199	Suzhou	CHN	65.88	248	Shenzhen	CHN	57.70
200	London	CAN	65.64	249	Kraków	POL	57.33
201	Liège	BEL	65.59	250	Panamá	PAN	57.32
202	Dortmund	DEU	65.46	251	Windhoek	NAM	57.15
203	Toulouse	FRA	65.22	252	Vilnius	LTU	56.49
204	València	ESP	64.46	253	Dali	CHN	56.30
205	Edmonton	CAN	64.36	254	Kobe	JPN	56.23
206	Taipei	TWN	64.33	255	Viña del Mar	CHL	55.85
207	Sjanghai	CHN	64.23	256	Surrey	CAN	55.04
208	Genova	ITA	64.21	257	Hiroshima	JPN	55.01
209	Mannheim	DEU	63.93	258	Zagreb	HRV	54.87
210	Lahore	PAK	63.48	259	Kaohsiung	TWN	54.85
211	Fez	MAR	63.36	260	Bratislava	SVK	54.77
212	Datong	CHN	63.30	261	Gelsenkirchen	DEU	54.35
213	Budapest	HUN	63.29	262	Ahmedabad	IND	54.11
214	Haifa	ISR	63.22	263	Istanbul	TUR	53.78
215	Nantes	FRA	63.21	264	Brno	CZE	53.69
216	Matsuyama	JPN	63.05	265	Tijuana	MEX	53.45
217	Doha	QAT	62.93	266	Shillong	IND	53.42
218	Zaragoza	ESP	62.45	267	Casablanca	MAR	53.18
219	Kitakyushu	JPN	62.31	268	Alicante	ESP	52.96
220	Rouen	FRA	62.24	269	Grenoble	FRA	52.93
221	Wenzhou	CHN	61.84	270	Toulon	FRA	52.83
222	Sanya	CHN	61.79	271	Zhaoqing	CHN	52.51
223	Mumbai	IND	61.70	272	Kampala	UGA	52.36
224	Agadir	MAR	61.69	273	Mérida	MEX	52.05
225	Stoke-on-Trent	GBR	61.60	274	Sharjah	ARE	52.02
226	Kumamoto	JPN	61.43	275	Bydgoszcz	POL	51.88
227	Abidjan	CIV	61.33	276	Colombo	LKA	51.80
228	Takamatsu	JPN	61.18	277	New Taipei	TWN	51.60
229	Katowice	POL	61.11	278	Taichung	TWN	51.59
230	Bochum	DEU	60.88	279	Gdansk	POL	51.33
231	Huancayo	PER	60.87	280	Wroclaw	POL	51.13
232	Essen	DEU	60.84	281	Douala	CMR	50.93
233	Kitchener	CAN	60.71	282	San Pedro Sula	HND	50.65
234	Bari	ITA	60.68	283	Kyiv	UKR	50.56
235	Petah Tikva	ISR	60.61	284	Murcia	ESP	50.51
236	Nagoya	JPN	60.54	285	Yangzhou	CHN	50.43
237	Saint Petersburg	RUS	59.93	286	Addis Ababa	ETH	50.42
238	Riga	LVA	59.83	287	Durban	ZAF	50.39
239	Shaoxing	CHN	59.66	288	Xiangyang	CHN	50.30
240	Las Palmas	ESP	59.37	289	Amman	JOR	49.90
241	Nairobi	KEN	59.26	290	Dalian	CHN	49.81
242	Vladivostok	RUS	59.12	291	Warszawa	POL	49.75

292	Valledupar	COL	49.54
293	Duisburg	DEU	49.32
294	Hsinchu	TWN	48.97
295	Poznan	POL	48.68
296	Guangzhou	CHN	48.65
297	Lublin	POL	48.44
298	Oaxaca	MEX	48.13
299	Palermo	ITA	47.98
300	Catania	ITA	47.92
301	Lusaka	ZMB	47.80
302	Santa Cruz d. Tenerife	ESP	47.70
303	Prayagraj	IND	47.57
304	Mazatlán	MEX	47.11
305	Johannesburg	ZAF	47.05
306	Kigali	RWA	46.71
307	Taoyuan	TWN	46.66
308	Muscat	OMN	46.60
309	Tianjin	CHN	46.59
310	Mombasa	KEN	46.56
311	Chiayi	TWN	46.47
312	Wuppertal	DEU	46.42
313	Karachi	PAK	46.33
314	Toshkent	UZB	45.94
315	Tangier	MAR	45.87
316	Liangshan	CHN	45.87
317	Changsha	CHN	45.76
318	Samara	RUS	45.63
319	Abuja	NGA	45.59
320	Quanzhou	CHN	45.40
321	George Town	MYS	45.32
322	Cotonou	BEN	45.30
323	Meknes	MAR	45.29
324	Binhai New Area	CHN	45.13
325	Huizhou	CHN	45.08
326	Nizhny Novgorod	RUS	45.05
327	Chengdu	CHN	45.00
328	Ciudad de México	MEX	44.85
329	Udaipur	IND	44.76
330	Bandaraya Melaka	MYS	44.71
331	Huaiyin	CHN	44.70
332	Daegu	KOR	44.49
333	Zhanjiang	CHN	44.46
334	Acapulco	MEX	44.40
335	Nanjing	CHN	43.98
336	Omsk	RUS	43.97
337	Thessaloniki	GRC	43.91
338	Fuzhou	CHN	43.86
339	Chongqing	CHN	43.86
340	Temuco	CHL	43.62
341	Ningbo	CHN	43.38
342	Buenos Aires	ARG	43.29
343	Minsk	BLR	43.13
344	Rabat	MAR	42.96
345	Port Elizabeth	ZAF	42.87
346	Kuala Lumpur	MYS	42.75
347	Kazan	RUS	42.49
348	Changzhou	CHN	42.48
349	Plovdiv	BGR	42.45
350	Rio de Janeiro	BRA	42.40
351	Harare	ZWE	42.37
352	Kagoshima	JPN	42.34
353	Nha Trang	VNM	42.25
354	Antalya	TUR	42.15
355	Huangdao District	CHN	42.11
356	Yaoundé	CMR	42.11
357	San José	CRI	41.90
358	Sandy	USA	41.82
359	Dhaka	BGD	41.62
360	Cuernavaca	MEX	41.50
361	Algiers	DZA	41.33
362	Weifang	CHN	41.19
363	Santo Domingo	DOM	40.73
364	Lodz	POL	40.72
365	West Valley City	USA	40.59
366	Chiang Mai	THA	40.58
367	Delhi	IND	40.50
368	Xining	CHN	40.34
369	Florianópolis	BRA	40.13
370	Monterrey	MEX	40.13
371	Tétouan	MAR	39.99
372	Port-au-Prince	HTI	39.85
373	Meilan District	CHN	39.85
374	Maputo	MOZ	39.83
375	Santiago	DOM	39.70
376	Nanchang	CHN	39.66
377	Wuhan	CHN	39.62
378	Tangshan	CHN	39.58
379	Ho Chi Minh City	VNM	39.58
380	Jaipur	IND	39.57
381	Guayaquil	ECU	39.50
382	Mexicali	MEX	39.39
383	Santa Marta	COL	39.37
384	Bucharest	ROU	39.34
385	Zhenjiang	CHN	39.14
386	São Paulo	BRA	39.10
387	Concepción	CHL	39.03
388	Ürümqi	CHN	38.82
389	Cancún	MEX	38.78

390	Busan	KOR	38.69	439	Rostov-on-Don	RUS	35.07
391	Yekaterinburg	RUS	38.65	440	Santa Fe	ARG	34.88
392	Weihai	CHN	38.56	441	Mendoza	ARG	34.84
393	Meizhou	CHN	38.48	442	Jiaxing	CHN	34.82
394	Morelia	MEX	38.42	443	Kenitra	MAR	34.74
395	Baguio	PHL	38.38	444	Sarajevo	BIH	34.63
396	Montevideo	URY	38.27	445	Corrientes	ARG	34.63
397	Yerevan	ARM	38.05	446	Yantai	CHN	34.62
398	Nanning	CHN	38.05	447	Tolyatti	RUS	34.61
399	Wuxi	CHN	38.04	448	Coimbatore	IND	34.60
400	Shijiazhuang	CHN	37.60	449	Rawalpindi	PAK	34.53
401	La Habana	CUB	37.55	450	Lviv	UKR	34.48
402	Gwangju	KOR	37.53	451	Orizaba	MEX	34.20
403	Freetown	SLE	37.47	452	Qinhuangdao	CHN	34.18
404	Santa Cruz d. l. Sierra	BOL	37.40	453	Tver	RUS	34.09
405	Phnom Penh	KHM	37.38	454	Santiago	CHL	34.08
406	Guilin	CHN	37.27	455	Oran	DZA	34.00
407	Lanzhou	CHN	37.27	456	Salta	ARG	33.88
408	Nur-Sultan	KAZ	37.23	457	Makassar	IDN	33.82
409	Antananarivo	MDG	37.14	458	Puebla	MEX	33.76
410	Tampico	MEX	37.13	459	Nangang	CHN	33.70
411	Pukou	CHN	36.99	460	Subang Jaya	MYS	33.68
412	Kandy	LKA	36.80	461	Baku	AZE	33.68
413	Xi'an	CHN	36.72	462	Pohang-si	KOR	33.47
414	Querétaro	MEX	36.69	463	Bengaluru	IND	33.37
415	Daejeon	KOR	36.63	464	Ipoh	MYS	33.37
416	Zhongshan	CHN	36.63	465	Krasnodar	RUS	33.31
417	Cairo	EGY	36.62	466	Izhevsk	RUS	33.31
418	Belgrade	SRB	36.58	467	Cúcuta	COL	33.27
419	Dakar	SEN	36.53	468	Pucallpa	PER	32.86
420	Santos	BRA	36.45	469	Chihuahua	MEX	32.84
421	Rizhao	CHN	36.42	470	Piura	PER	32.84
422	Odesa	UKR	36.16	471	Salvador	BRA	32.67
423	Quanshan	CHN	36.14	472	Tbilisi	GEO	32.66
424	Pune	IND	35.97	473	Liuzhou	CHN	32.56
425	Kota Kinabalu	MYS	35.92	474	Novokuznetsk	RUS	32.55
426	Zanzibar City	TZA	35.88	475	Ciudad Obregón	MEX	32.47
427	Hrodna	BLR	35.82	476	Ulsan	KOR	32.44
428	Kochi	IND	35.80	477	Volgograd	RUS	32.31
429	Maceió	BRA	35.77	478	Celaya	MEX	32.28
430	Sofia	BGR	35.75	479	Darjeeling	IND	32.28
431	Zhuhai	CHN	35.72	480	Barranquilla	COL	32.23
432	Chisinau	MDA	35.45	481	Jingdezhen	CHN	32.20
433	Yichang	CHN	35.44	482	Luoyang	CHN	32.17
434	Praia Grande	BRA	35.44	483	Tula	RUS	32.16
435	Varanasi	IND	35.36	484	Dnipro	UKR	32.04
436	Mar del Plata	ARG	35.33	485	Puducherry	IND	32.03
437	Guiyang	CHN	35.15	486	Dar es-Salaam	TZA	31.99
438	Pretoria	ZAF	35.10	487	Honghuagang	CHN	31.94

488	Vung Tau	VNM	31.81	537	Davao City	PHL	29.55
489	Hanoi	VNM	31.79	538	Sousse	TUN	29.50
490	Astrakhan	RUS	31.77	539	Tiexi	CHN	29.43
491	Posadas	ARG	31.67	540	Ciudad de Guatemala	GTM	29.41
492	Jinan	CHN	31.62	541	Changchun	CHN	29.25
493	Zhangzhou	CHN	31.60	542	Yaroslavl	RUS	29.22
494	Alajuela	CRI	31.53	543	Campinas	BRA	29.14
495	Tyumen	RUS	31.51	544	Lipetsk	RUS	29.10
496	Quito	ECU	31.50	545	Dongguan	CHN	29.08
497	Leshan	CHN	31.44	546	São José do Rio Preto	BRA	29.06
498	Tegucigalpa	HND	31.40	547	Novosibirsk	RUS	28.99
499	Fortaleza	BRA	31.38	548	Wuhu	CHN	28.99
500	Zhuzhou	CHN	31.37	549	Taiyuan	CHN	28.91
501	Baotou	CHN	31.35	550	Medellín	COL	28.87
502	Jiujiang	CHN	31.24	551	Campo Grande	BRA	28.84
503	Phuket	THA	31.24	552	Kaliningrad	RUS	28.81
504	Xishan	CHN	31.21	553	Eskisehir	TUR	28.80
505	Kaifeng	CHN	31.21	554	Cuiabá	BRA	28.67
506	Cusco	PER	31.11	555	Taguatinga	BRA	28.64
507	Nantong	CHN	31.06	556	La Plata	ARG	28.53
508	Almaty	KAZ	31.06	557	Panlong	CHN	28.52
509	Hefei	CHN	31.05	558	Mianyang	CHN	28.50
510	Ciudad Juárez	MEX	31.02	559	Irapuato	MEX	28.39
511	Magnitogorsk	RUS	30.98	560	Asunción	PRY	28.33
512	Cagayan de Oro	PHL	30.95	561	Toluca	MEX	28.13
513	Dujiangyan	CHN	30.90	562	Foshan	CHN	28.12
514	Chennai	IND	30.84	563	Skopje	MKD	28.11
515	San Juan	ARG	30.80	564	Bacolod	PHL	28.11
516	Rosario	ARG	30.80	565	Khabarovsk	RUS	28.10
517	Serrekunda	GMB	30.72	566	Goiânia	BRA	27.97
518	Joinville	BRA	30.71	567	Izmir	TUR	27.90
519	Manila	PHL	30.70	568	Belém	BRA	27.85
520	Tirana	ALB	30.67	569	Curitiba	BRA	27.84
521	Aguascalientes	MEX	30.62	570	Bukit Mertajam	MYS	27.82
522	Jodhpur	IND	30.62	571	San Luis Potosí	MEX	27.74
523	Kota Bharu	MYS	30.58	572	Chandigarh	IND	27.69
524	Beihai	CHN	30.55	573	Canoas	BRA	27.66
525	Penza	RUS	30.46	574	Niterói	BRA	27.66
526	Jinhua	CHN	30.28	575	San Salvador	SLV	27.65
527	Guadalajara	MEX	30.21	576	Medan	IDN	27.61
528	Córdoba	ARG	30.20	577	Zibo	CHN	27.60
529	Ribeirão Preto	BRA	30.12	578	Caxias do Sul	BRA	27.51
530	Vila Velha	BRA	30.06	579	Lomé	TGO	27.41
531	Tunis	TUN	29.98	580	Taishan District	CHN	27.39
532	Huadu	CHN	29.90	581	Vinnytsia	UKR	27.36
533	Cuenca	ECU	29.76	582	Yangon	MMR	27.26
534	Hyderabad	IND	29.63	583	Pereira	COL	27.23
535	Zigong	CHN	29.56	584	Belo Horizonte	BRA	27.23
536	Hohhot	CHN	29.56	585	Kajang	MYS	27.17

586	Mysuru	IND	27.16
587	João Pessoa	BRA	27.14
588	Yanji	CHN	27.06
589	Brest	BLR	27.03
590	Vientiane	LAO	26.92
591	Qingyuan	CHN	26.88
592	Villahermosa	MEX	26.81
593	Cheonan-si	KOR	26.78
594	Bursa	TUR	26.77
595	Bishkek	KGZ	26.77
596	Mahilyow	BLR	26.75
597	Zhengzhou	CHN	26.71
598	Santiago de Cali	COL	26.71
599	Belgorod	RUS	26.69
600	Bogotá	COL	26.67
601	Mangaluru	IND	26.67
602	Chiclayo	PER	26.65
603	Recife	BRA	26.62
604	Zhangjiakou	CHN	26.53
605	León	MEX	26.51
606	Pikine	SEN	26.47
607	Teresina	BRA	26.45
608	Porto Alegre	BRA	26.40
609	Can Tho	VNM	26.39
610	Pachuca	MEX	26.34
611	Baoding	CHN	26.21
612	Torreón	MEX	26.19
613	Ulaanbaatar	MNG	26.10
614	Jaboatão dos Guara.	BRA	26.00
615	Cebu City	PHL	25.98
616	Xinxiang	CHN	25.98
617	Bhubaneshwar	IND	25.97
618	Kharkiv	UKR	25.95
619	Irkutsk	RUS	25.94
620	Jundiaí	BRA	25.90
621	Quezon City	PHL	25.90
622	Iloilo City	PHL	25.80
623	Shantou	CHN	25.77
624	Yogyakarta	IDN	25.69
625	Tomsk	RUS	25.63
626	Durango	MEX	25.48
627	Uberlândia	BRA	25.38
628	Voronezh	RUS	25.29
629	Dandong	CHN	25.29
630	Hengyang	CHN	25.26
631	Manaus	BRA	25.22
632	Ibagué	COL	25.21
633	Surakarta	IDN	25.17
634	Beibei	CHN	25.17
635	Krasnoyarsk	RUS	25.05
636	Hermosillo	MEX	25.03
637	Indore	IND	25.02
638	Vitsebsk	BLR	25.00
639	Villavicencio	COL	24.93
640	Naberezhnye Chelny	RUS	24.88
641	Veracruz	MEX	24.80
642	Natal	BRA	24.65
643	Juiz de Fora	BRA	24.64
644	Kuching	MYS	24.57
645	Coatzacoalcos	MEX	24.55
646	Londrina	BRA	24.53
647	Saltillo	MEX	24.40
648	Da Nang	VNM	24.35
649	Yinchuan	CHN	24.33
650	Saratov	RUS	24.31
651	Culiacán	MEX	24.23
652	Seremban	MYS	23.99
653	Smolensk	RUS	23.76
654	Tucumán	ARG	23.65
655	São Luís	BRA	23.60
656	Kathmandu	NPL	23.45
657	Jingzhou	CHN	23.42
658	Cheboksary	RUS	23.32
659	Hengshui	CHN	23.27
660	Aracaju	BRA	23.24
661	Yongchuan	CHN	23.08
662	Arequipa	PER	23.05
663	Changping	CHN	22.85
664	Ufa	RUS	22.83
665	Langfang	CHN	22.81
666	Chelyabinsk	RUS	22.59
667	Klang	MYS	22.54
668	Hue	VNM	22.49
669	Ivanovo	RUS	22.40
670	São José dos Campos	BRA	22.38
671	Hunnan	CHN	22.31
672	Quilpué	CHL	22.25
673	Homyel	BLR	22.18
674	Maringá	BRA	22.05
675	Dehradun	IND	21.99
676	Novo Hamburgo	BRA	21.98
677	Bauru	BRA	21.97
678	Arusha	TZA	21.80
679	Agra	IND	21.71
680	Jilin	CHN	21.63
681	Mykolaiv	UKR	21.27
682	Ulan-Ude	RUS	21.22
683	Jakarta	IDN	21.19

684	Lima	PER	21.09
685	Ryazan	RUS	21.06
686	Ouagadougou	BFA	20.87
687	Kemerovo	RUS	20.84
688	La Paz	BOL	20.81
689	Uberaba	BRA	20.77
690	Batam	IDN	20.75
691	Ankara	TUR	20.72
692	Santiago de Cuba	CUB	20.69
693	Tepic	MEX	20.41
694	Riyadh	SAU	20.40
695	Semarang	IDN	20.15
696	Neiva	COL	20.07
697	Campina Grande	BRA	20.03
698	Pasto	COL	20.02
699	Mbour	SEN	20.01
700	Nagpur	IND	19.79
701	Shunyi	CHN	19.68
702	Stavropol	RUS	19.65
703	Mataram	IDN	19.53
704	Angeles	PHL	19.33
705	Orenburg	RUS	19.21
706	Cochabamba	BOL	19.11
707	Nashik	IND	19.00
708	Zaporizhzhia	UKR	18.95
709	Xalapa	MEX	18.62
710	Kisumu	KEN	18.37
711	Sorocaba	BRA	18.11
712	Tuxtla Gutiérrez	MEX	17.94
713	Manizales	COL	17.73
714	Guwahati	IND	17.58
715	Bucaramanga	COL	17.52
716	Barnaul	RUS	17.42
717	Luxor	EGY	16.79
718	Vadodara	IND	15.96
719	Lianyungang	CHN	15.72
720	Surabaya	IDN	14.53
721	Piracicaba	BRA	14.45
722	Trujillo	PER	13.99
723	Lucknow	IND	13.31
724	Villa Nueva	GTM	13.27
725	Bryansk	RUS	13.11
726	San Lorenzo	PRY	12.95
727	Managua	NIC	12.78
728	Bandung	IDN	10.95
729	Kumasi	GHA	10.79
730	Jeddah	SAU	10.72
731	Montería	COL	10.44
732	Srinagar	IND	10.28
733	Caracas	VEN	9.82
734	Mandalay	MMR	5.54

Table A2: Correlations, price of a representative apartment at the city center

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	1.00												
B	0.96	1.00											
C	0.99	0.96	1.00										
D	0.99	0.96	0.99	1.00									
E	1.00	0.96	0.99	0.99	1.00								
F	0.97	0.94	0.97	0.96	0.97	1.00							
G	1.00	0.96	0.99	0.99	0.99	0.96	1.00						
H	1.00	0.96	0.99	0.99	1.00	0.97	1.00	1.00					
J	0.98	0.95	0.98	0.97	0.98	0.96	0.98	0.98	1.00				
K	0.95	0.93	0.95	0.94	0.95	0.93	0.95	0.95	0.97	1.00			
L	0.98	0.96	0.98	0.99	0.98	0.96	0.98	0.98	0.99	0.96	1.00		
M	1.00	0.96	0.99	0.99	0.99	0.97	0.99	1.00	0.97	0.95	0.98	1.00	
N	0.92	0.90	0.92	0.92	0.91	0.92	0.91	0.92	0.95	0.94	0.95	0.91	1.00

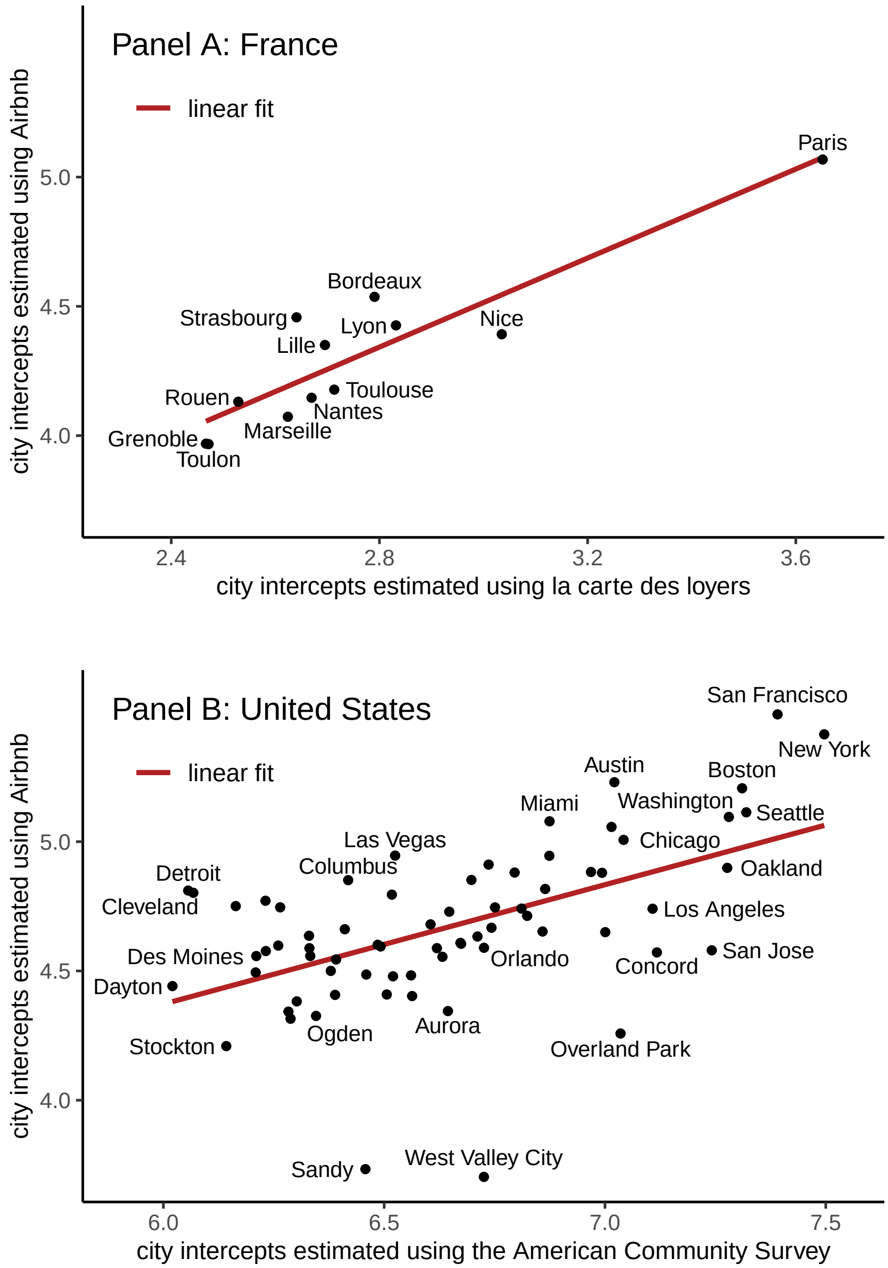
Note: For this table, I recompute the ranking of cities by their estimated nightly short-term rental rate of a representative apartment at the city center (Table A1). The table reports correlation coefficients of the estimated USD prices among all rankings. Specification A refers to the baseline version as shown in Table A1. B is based only on entire apartments, excluding properties that are shared. C controls for ratings. D controls for all possible amenities. E does not control for proximity to the shore of an ocean or big lake. F includes all properties that have been rented at least once. G includes all properties that have been rented or available at least 125 of 365 days. H uses prices not windsorized to the 0.01 and 0.99 percentiles by country. Specification I uses the centers from OSM for all cities. J uses the centers from Google Maps for all cities. K uses ln(distance) instead of ln(distance + 1) to compute the distance gradients. L divides the properties in each city in halfs, according to their distance to the city center, and then demeans by city halfs. Finally, M uses average retal prices instead of rental prices at the city center by not including distance gradients.

Table A3: Rank correlations, price of a representative apartment at the city center

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	1.00												
B	0.94	1.00											
C	0.98	0.94	1.00										
D	0.99	0.94	0.97	1.00									
E	1.00	0.94	0.98	0.98	1.00								
F	0.93	0.89	0.93	0.92	0.93	1.00							
G	0.99	0.93	0.97	0.98	0.99	0.92	1.00						
H	1.00	0.94	0.98	0.99	1.00	0.93	0.99	1.00					
J	0.96	0.91	0.95	0.95	0.96	0.93	0.95	0.96	1.00				
K	0.92	0.89	0.92	0.91	0.92	0.90	0.91	0.92	0.95	1.00			
L	0.96	0.93	0.95	0.98	0.96	0.92	0.96	0.96	0.97	0.94	1.00		
M	0.99	0.93	0.98	0.98	0.99	0.93	0.99	0.99	0.95	0.92	0.95	1.00	
N	0.89	0.88	0.89	0.90	0.88	0.89	0.88	0.89	0.94	0.92	0.95	0.88	1.00

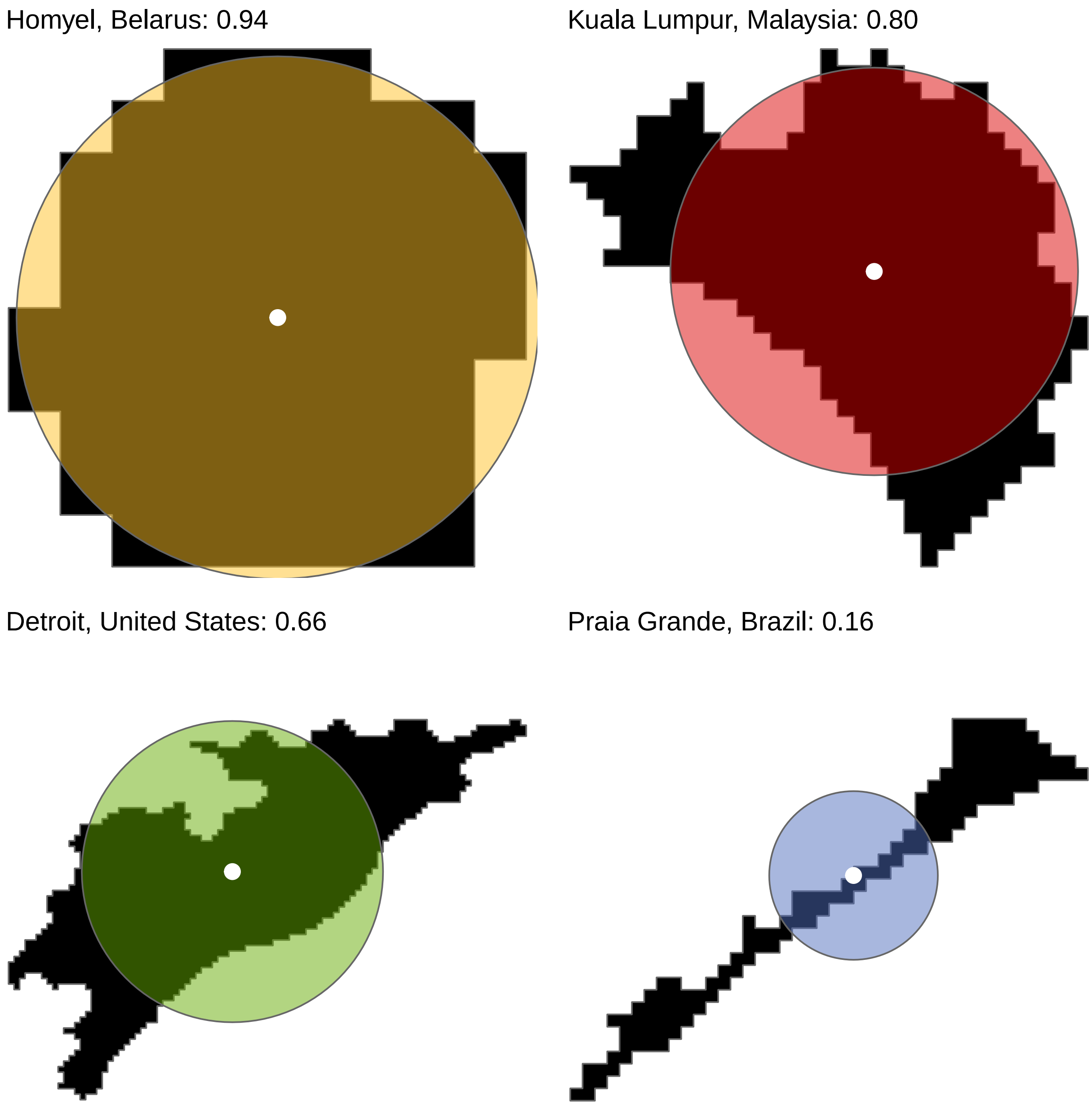
Note: For this table, I recompute the ranking of cities by their estimated nightly short-term rental rate of a representative apartment at the city center (Table A1). The table reports correlation coefficients of the estimated rank positions among all rankings. Specification A refers to the baseline version as shown in Table A1. B is based only on entire apartments, excluding properties that are shared. C controls for ratings. D controls for all possible amenities. E does not control for proximity to the shore of an ocean or big lake. F includes all properties that have been rented at least once. G includes all properties that have been rented or available at least 125 of 365 days. H uses prices not windsorized to the 0.01 and 0.99 percentiles by country. Specification I uses the centers from OSM for all cities. J uses the centers from Google Maps for all cities. K uses ln(distance) instead of ln(distance + 1) to compute the distance gradients. L divides the properties in each city in halfs, according to their distance to the city center, and then demeans by city halfs. Finally, M uses average retal prices instead of rental prices at the city center by not including distance gradients.

Figure A3: Comparison



Note: This figure shows correlations between city intercepts estimated from short-term and long-term rental data. Panel A shows the analysis for France, using commune-level rents from *la carte des loyers*, who estimate them as outputs of hedonic regressions. I match 451 communes (and arrondissements) to my 12 (functional) French cities. I then regress prices on city distance gradients and city intercepts. The x-axis of Panel A shows these estimated intercepts. Panel B shows the analysis for the United States, using block group level rents from the 2015-2019 American community survey, which are averages of survey responses. I match 43,636 block groups to my (functional) US cities. I then regress prices on city distance gradients and city intercepts. In this case, I control for the block groups' fractions of several building-related characteristics. The x-axis of Panel B shows these estimates. In both cases, the y-axis shows the city intercepts that are the outcome of my first-stage regression using properties from Airbnb that lead to the ranking in Table A1.

Figure A4: Compactness



Note: The figure shows the compactness measure that I use as a second-stage control variable for four exemplary cities. The chosen cities have compactness measures that correspond to the maximal value, 75% quantile, 25% quantile, and minimal value of the distribution. The measure is taken from Angel et al. (2020) where it is called "exchange". To create it, I compute a circle with the same area as the city itself around each city's centroid and then measure the proportion of the circle that intersects with the shape of the city.

Table A4: Additional specifications

Dependent Variable:	log(Price of representative apartment at city center)					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
log(Population)	-0.435***		-0.330***		0.164***	0.161***
	(0.062)		(0.083)		(0.057)	(0.050)
log(Area)	0.639***		0.516***		-0.030	-0.050
	(0.068)		(0.104)		(0.063)	(0.053)
Compactness		-0.198	0.078			-0.054
		(0.313)	(0.195)			(0.112)
Elevation (100m)		-0.014***	-0.004			-0.007*
		(0.003)	(0.004)			(0.004)
Difference to 21.11°C		0.010	0.004			0.002
		(0.014)	(0.012)			(0.007)
By ocean / big lake		0.111**	0.100***			0.054
		(0.051)	(0.035)			(0.045)
Capital		0.123	0.114**			0.105*
		(0.100)	(0.055)			(0.057)
Airbnbs per 1,000		0.058***	0.048***			0.029***
		(0.010)	(0.009)			(0.006)
Country fixed effects	-	-	-	Yes	Yes	Yes
R^2	0.325	0.281	0.475	0.736	0.776	0.813

Note: The table shows regressions of the estimated price of a representative short-term rental property at the city center on city size and control variables. The units of observation are 733 cities. Column 4 includes only country fixed effects. The parentheses show standard errors, which are clustered by country. The levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Heterogeneity by country without the countries' two largest cities

Dependent Variable:	log(Price of representative apartment at city center)						
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	USA	Eurozone	Russia	China	India	Brazil	Mexico
log(Population)	0.23**	0.30***	0.02	0.10	0.42*	0.04	-0.45**
	(0.11)	(0.09)	(0.26)	(0.12)	(0.22)	(0.09)	(0.19)
log(Area)	-0.07	-0.10	0.14	0.01	-0.23	0.04	0.59***
	(0.12)	(0.09)	(0.26)	(0.13)	(0.23)	(0.11)	(0.19)
Compactness	-0.36**	0.35	-0.34	-0.34	-0.17	0.34	-0.65
	(0.18)	(0.36)	(0.43)	(0.32)	(1.30)	(0.24)	(0.54)
Elevation (100m)	-0.02***	-0.01	-0.04*	0.00	0.00	0.00	0.00
	(0.01)	(0.03)	(0.02)	(0.01)	(0.03)	(0.01)	(0.01)
Difference to 21.11°C	0.02**	0.08***	0.00	-0.01**	-0.05	0.00	-0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.05)	(0.01)	(0.02)
By ocean / big lake	-0.11**	-0.03	0.04	0.22***	0.15	0.01	0.07
	(0.05)	(0.07)	(0.07)	(0.08)	(0.29)	(0.07)	(0.14)
Airbnbs per 1,000	0.04***	0.02***	0.06**	0.03**	0.78**	0.10***	0.10***
	(0.01)	(0.00)	(0.03)	(0.01)	(0.36)	(0.03)	(0.02)
Country fixed effects	-	Yes	-	-	-	-	-
Observations	68	74	42	110	29	42	36

Note: The table shows regressions of the estimated price of a representative short-term rental property at the city center on city size and control variables. The units of observation are cities. The two cities with the largest population in each entity are excluded. The parentheses show standard errors clustered by country for specification 2 (eurozone) and heteroscedasticiy robust standard errors for all other specifications. The levels of significance are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Comparison with Chauvin et al. (2017)

		USA		Brazil	China		India
		ln(rent)	ln(price)	ln(rent)	ln(rent)	ln(price)	ln(rent)
Chauvin et al. (2017)	OLS	0.15***	0.20***	0.13***	0.23***	0.10	0.003
		(0.01)	(0.04)	(0.02)	(0.08)	(0.12)	(0.005)
	IV	0.15***	0.20***	0.13***	0.37***	0.06	−0.004
		(0.01)	(0.04)	(0.02)	(0.13)	(0.13)	(0.009)
		log(nightly rate of representative apartment at city center)					
Preferred	OLS	0.17***		0.09***	0.13***		0.23***
		(0.03)		(0.02)	(0.03)		(0.08)
	IV	0.18***		0.08***	0.13***		0.23**
		(0.03)		(0.02)	(0.03)		(0.09)
No controls	OLS	0.19***		0.11***	0.14***		0.16*
		(0.04)		(0.03)	(0.03)		(0.08)
	IV	0.20***		0.10***	0.11***		0.16*
		(0.04)		(0.03)	(0.04)		(0.08)
		log(nightly rate of representative apartment anywhere in the city)					
No controls, means	OLS	0.12***		0.09***	0.06**		0.03
		(0.03)		(0.03)	(0.02)		(0.04)
	IV	0.12***		0.10***	0.02		0.02
		(0.03)		(0.03)	(0.03)		(0.04)

Note: The first strip shows the estimates Chauvin et al. (2017) obtain when regressing either log rents or log house prices on log population. They measure population in 2010 and use population in 1980 as an instrument. The other three stripes are based on my own estimates. The dependent variable of the second and the third strip is the estimated price of a representative short-term rental property at the city center. The dependent variable of the fourth strip is the estimated price of a representative short-term rental property anywhere in the city. The difference between the two is whether the first-stage hedonic regression does (strip 2 and 3) or does not (strip 4) include distance gradients. The second strip is estimated using all control variables that are included in Table 3, but does not control for area. The third and the fourth strip are estimated using neither controls nor controlling for area. Population in 2015 is instrumented by population in 1975 for the IV specifications of stripes 2, 3, and 4. The parentheses show heteroscedasticity robust standard errors. The levels of significance are * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.