Housing Supply in the Presence of Informality*

Ricardo Guedes,[†] Felipe S. Iachan,[‡] and Marcelo Sant'Anna[§]

January 16, 2023

Abstract

Informal housing refers to the unauthorized residential use of public or private land. In this paper, we study housing supply in markets where informal housing is common. Using a combination of census and satellite data, we estimate housing supply for more than 90 metropolitan areas in Brazil. We find that widespread informal housing increases the housing supply elasticity, partially offsetting the downward pressure of geographical constraints. Our empirical approach is guided by a monocentric city model that includes informal housing. Our identification strategy relies on the use of two novel instruments, combining demographic data and public land ownership.

Keywords: Housing Supply, Real Estate Markets, Informality, Slums

JEL: R31, O17, O18

*We are grateful to Fernando Ferreira, Daniel da Mata, Cecilia Machado, Ciro Biderman, and participants at the SBE 2020, LACEA-LAMES 2019, and European UEA 2021 meetings for helpful comments. Financial support from Rede de Pesquisa Aplicada FGV and from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001- is gratefully acknowledged. Declarations of interest: none.

†FGV EPGE. e-mail: ricardobg@gmail.com

‡FGV EPGE. e-mail: felipe.iachan@fgv.br

§Corresponding author. FGV EPGE. e-mail: marcelo.santanna@fgv.br. Address: Praia de Botafogo 190/1100, Rio de Janeiro RJ, Brazil.

Housing supply elasticity is a key factor for understanding urban dynamics. It influences how city growth, migration, and housing prices respond, for instance, to increased productivity and credit expansions. The literature on developed countries has focused on the central role of geographic constraints and regulation in shaping the supply elasticity.[1] For instance, American cities with less available land face more inelastic supply, producing stronger price responses to demand shocks (Saiz, 2010).

In this paper, we study a middle-income country setting in which informal land use plays a crucial role in the production of housing. Informality in the housing market, defined here as the unauthorized residential use of public or private land, is a major issue in most of the developing world. Informal housing raises many concerns, especially due to the lack of basic urban infrastructure that torments slum dwellers. In this paper, we focus on the effects of slums on the total housing supply (including formal and informal settlements). Because informal settlements develop without regard to any land use regulations, it is possible that they facilitate growth and thus contribute to increasing the housing supply elasticity. Additionally, if slums are more widespread in cities with stricter geographical constraints, their presence can confound the effect of land availability. We investigate these relations using data from metropolitan areas in Brazil. This application is guided by an otherwise standard monocentric city model, which we extend to incorporate the informal use of land.

In our model, informal housing does not compete with the formal market for land. Instead, informal land use occurs in areas that have been left vacant by formal urbanization: it originates in either public or private dormant land, with the formation of large illegal settlements. The lack of formal ownership always implies a risk of eviction. In practice, some types of land may be more prone to squatting, as the political and legal costs of removal may vary. For instance, it is typically considered more difficult to displace a settlement occupying public land. To incorporate the drawbacks of these informal living arrangements, we include two penalties from informal housing: an expected cost from the loss of physical assets in eviction and a flow penalty from the lack of proper urban infrastructure.

Households prefer to live near the city center to save time and money commuting. Informal housing dwellers, in particular, are willing to accept lower quality housing to live closer to the workplace. Rents adjust in equilibrium in the city, such that the homogeneous households are indifferent between locations and induced quality of both formal

[1] This connection has been exploited to study a range of problems, from location choice (e.g., Gyourko et al., 2013; Diamond, 2016, 2017) to the role of housing collateral on credit (e.g., Mian and Sufi, 2011; Chaney et al., 2012; Robb and Robinson, 2014).

and informal dwellings. Our model nests the model in Saiz (2010) as a particular case in which squatting is not possible, so for cities that do not feature slums the housing supply curve will be mainly shaped by geographical constraints. We therefore focus our empirical housing supply analysis on the set of metropolitan areas that feature slums (49% of metropolitan areas), in which the effects of informality in the housing supply are salient. We thus seek to understand how the level of informality together with geographical constraints shape the housing supply of theses cities.

There are a few important lessons from our model that guide our empirical exercise. First, we show that informality is decreasing in the amount of developable land. That is, cities that face more stringent geographical restrictions are predicted to have a higher share of informal housing. The intuition is that cities with more restrictions "push" its citizens farther out, increasing transportation costs. This creates incentives for informal housing close to the city's center. This theoretical relation, which we empirically confirm, suggests that informal housing may indeed confound the effect of geographical constraints in the housing supply.

The model establishes that city size, as measured by the number of households, is positively related to the share of households living in informal housing. This poses a challenge for any empirical investigation of the housing supply curve, as both city size and the share households in informal housing are subject to endogeneity stemming from supply and demand simultaneity. However, our model also offers a solution for the estimation of the supply curve: we show that informal housing decreases with the eviction rate and increases with the share of land that is laid dormant in the process of formal urbanization. Therefore, we can use as an instrument for informal housing any observable exogenous variable that could be directly linked to eviction or dormant inner-city land.

Following this reasoning, we use an original instrument for informality: a measure of federally-owned land area in the city. First, the federal government in Brazil does not directly engage in urban policy, nor buys or sells any real estate without very significant hurdles. Federal property holdings have distant historical origins, supporting the instrument's plausible exogeneity. Additionally, evictions from federal property are particularly rare, suggesting the instrument's relevance. Indeed, we find that the federal real estate area in these cities is correlated with the informal occupation of land.

In our empirical investigation of the supply curve, we estimate the inverse supply curve. Therefore, we include total quantity (the number of households in the city) as a regressor. This is an endogenous regressor given supply and demand simultaneity. We also propose a (as far as we know) novel instrument. We calculate the rate of natural increase for the population older than 20 years in each city, considering expected deaths for each age-cohort. This rate of natural increase is a typical demand shifter and, thus, a good candidate for instrumenting quantities in the supply relationship.

We find that both geographical constraints and informal housing are important determinants of the housing supply elasticity. In specifications in which we include only geographical constraints, leaving informal housing out, we find that geographical constraints have an insignificant effect on the supply elasticity. This result shows that, in the Brazilian context, informal housing does indeed eclipse the effect of geographical constraints. Only when we include both geographical constraints and informal housing, we find geographical constraints have an effect in line with Saiz (2010) – i.e., less available land implies more inelastic housing supply curves.

Indeed, we find that more informal housing implies more elastic housing supply curves. This result is robust to a variety of specifications and confirms our original conjecture. We also show that the variation in informal housing across cities is roughly as important as geography in accounting for the dispersion in housing supply elasticity estimates. Decomposing our elasticity estimates for each city, we learn that geography alone would produce 50% of the dispersion in elasticity seen in the data.

Using the estimates from our main specification, we compute inverse housing supply elasticities for all the cities in our sample. We find that inverse elasticities vary from 0.02 to 0.99, which implies a wide range of responses to similar demand shocks. Our estimates imply a more elastic housing supply than what is found for the US (Saiz, 2010), which is consistent with the effect of widespread informality that we document. The average inverse elasticity and the effect of informality is in the ballpark to what Alves (2021) found for Brazil. In order to put these effects into perspective, we consider the predicted price increase for 2030 implied by the demand shift from demographic pressure since 2010. Our estimates project housing price increases ranging from 0% to 46%, a meaningful dispersion, as a consequence of predicted demographic growth that varies from 16% to 72%.

Our model also predicts which cities may have informal land use. For informal housing to emerge, a city needs enough sprawl so that high transportation costs can actually compensate the costs of living in informal housing near the city center. Both population and geographic constraints, for instance, make cities sprawl out and thus be more likely to develop slums. We take these predictions to the data and confirm them in a variety of specifications.

Brazil is a good place to investigate the relation between informal housing and housing supply given the large number of cities, with significant cross-sectional variation in informal settlements and geographical constraints. Housing and utility expenses represent a mean share of 20% of households total expenses in Brazil, which can be higher in large cities (IBGE, 2011). Affordability is a major reason for the proliferation of illegal housing in urban areas. There are more than 11 million people living in urban slums in Brazil, most of them in large metropolitan areas, adding up to more than 3 million houses.[2] Slum expansion meets a part of the demand for well-located homes, and is encouraged by the lack of available land, the availability of vacant urban property, the labor market, and resident's financial situation.

Related Literature– Previous studies examine the housing supply elasticity, the role of geographical constraints, and the role of informal markets. However, to the best of our knowledge, no research has particularly focused on the interaction of all three issues.

For example, some studies estimate the housing supply elasticity for countries other than the United States and focus on the consequences of geographical constraints, without a particular concern for informality (Hilber and Vermeulen, 2015; Malpezzi and Maclennan, 2001; Wang et al., 2012; Oikarinena et al., 2015; Harari, 2020). In particular, Combes et al. (2019) also use a monocentric city model with homogenous agents as a basis to study how housing costs increase with population in France.

Our paper contributes to a wide literature discussing the role of slums in urban development in developing countries. For instance, Celhay and Undurranga (2019) study a panel of housing decisions in Chile and provide evidence of a location-quality tradeoff involving better quality formal settlements that are distant from the city center and inferior quality informal settlements that are closer.

On the related theoretical literature, even though there is a established branch studying competition for land between formal households and squatters as determinants of house prices (e.g., Jimenez, 1985; Brueckner and Selod, 2009; Brueckner, 2013),[3] there are only a few empirical papers that incorporate slums as part of the housing supply in developing countries. For instance, Cai et al. (2018); Cavalcanti et al. (2019); Henderson et al. (2021) build models of city development in which slums play a crucial role in the housing supply but are estimated, or calibrated, for single cities.[4] Here we are interested in exploring how different patterns on urban land use and geographical determinants ex-

[2]In Brazil, there is a common misconception that slums occupy hills, but our data show this is not a predominant feature of informal settlements and most of them are located on flat land (i.e., under 15% slope). Slum dwellers construct their houses at lower costs near city centers. For more information on the determinants of slum formation and its relation to demographic and geographical characteristics, see Guedes (2020, chap. 2).

[3]While we abstract from land use competition, we focus on price competition and its effect on shaping the long-run housing price elasticity.

[4]There is also some, more distantly related, literature on the interactions between slum growth and developing-country issues, such as rural-urban migration, labor markets, and land-use policies (e.g., Marx et al., 2013; Ferreira et al., 2016; Da Mata et al., 2008; Smolka and Biderman, 2011).

plain the variation in housing supply elasticities in the cross-section of cities for a large developing country. In this respect, our paper relates to Alves (2021), that develops a model of slum growth and estimates separate supply elasticities for serviced and unserviced houses in Brazil, but abstracts from the interplay with geographical constraints.

In fact, some previous research finds correlation between illegal settlements and geographical constraints in Brazil, without a particular focus on overall housing market conditions (Nadalin and Mation, 2018). In this paper we provide a novel mechanism for this correlation in which tighter geographical constrains imply more sprawl, which in turn creates incentives for informal settlements.

1 Model

Our model is based on the monocentric city Alonso–Muth–Mills models (Alonso, 1964; Mills, 1967; Muth, 1969), unified by Brueckner (1987). We extend this framework to include geographical constraints (as in Saiz, 2010) and informal occupation of land by slums.

Households in city k are homogeneous and derive utility from city amenities (A_k) and consumption of a *numéraire* good.[5] Moreover, there is a penalty in (consumption-equivalent) utility from the distance (d) to the central business district (CBD) and from living in informal housing (Ψ_k):

$$U(C_k) = C_k^{\sigma} = \left(A_k + w_k - r' - td - \mathbb{1}_j \Psi_k\right)^{\sigma}, \tag{1}$$

where w_k is city wage level, r' is rent, t is commuting cost per unit of distance to CBD, and $\mathbb{1}_j$ is a dummy variable, which equals 1 if the household lives in an informal area, $j = i$, and zero if they live in a formal area, $j = f$. Informal land use takes place in spaces left vacant in the process of formal urban development. We choose the unit of distance, and therefore of the unit of area, so that one household lives on exactly one unit of area. We next focus on the relation between rents across sectors and distance to the CBD.

Because households are homogeneous, equilibrium in city location choices dictates that rents must adjust so that all households are indifferent between any type of dwelling, be it formal or informal, and location. This implies that rents throughout the city will be given by:

[5] We abstract from household heterogeneity in our model. There are two main reasons for this. First, data limitations prevent a separate empirical treatment for the formal and informal sectors (see Section 2). Second, although household heterogeneity might be an important driver of the demand for informal housing (e.g., Cavalcanti et al., 2019; Alves, 2021), our ultimate goal is to estimate the housing supply elasticity. We seek, thus, to provide a parsimonious model to guide this exercise.

$$r_{j,k}(d) = r_{f,k}(0) - td - \mathbb{1}_j \Psi_k, \text{ for } j \in \{i, f\}, \tag{2}$$

where $r_{f,k}(0)$ is formal rent in the CBD and $\mathbb{1}_j$ is the indicator for informality. This implies that rent differences across the two sectors at each distance must equal the disutility from living in informal housing, Ψ_k.

Housing supply. In a competitive market, developers break-even and the price of a new housing unit must equate to construction and land costs, as follows:

$$P_{j,k}(d) = CC + LC_{j,k}(d). \tag{3}$$

Importantly, land costs may differ between the formal and informal housing sectors, while it is not essential for our purposes that construction costs differ.

Equilibrium in the formal housing market implies that rents must exactly compensate the flow return of property value:

$$P_{f,k}(d) = \frac{r_{f,k}(d)}{\rho}, \tag{4}$$

where ρ denotes the interest rate.

We now assume one important difference between informal and formal housing. Informal housing dwellers are subject to a random chance of eviction at a yearly hazard rate λ_k.[6] Therefore, in the informal housing sector:

$$P_{i,k}(d) = \frac{r_{i,k}(d)}{\rho + \lambda_k}. \tag{5}$$

Let $\Phi_{f,k}$ and $\Phi_{i,k}$ denote the radii of formal and informal land use in the city. Land costs at the edge of the city's formal and informal housing developments must be zero. This ties rents to construction costs, interest rate, and eviction rate:

$$r_{f,k}(\Phi_{f,k}) = \rho \cdot CC, \tag{6}$$

$$r_{i,k}(\Phi_{i,k}) = (\rho + \lambda_k) \cdot CC. \tag{7}$$

We can now use equations (6) and (7) together with equation (2) to determine rents at

[6] Cavalcanti et al. (2019) also model the costs of living in slums as both an amenity shift and as separate cost associated with insecure land tenure, in their case involving "guard" labor. In our setting with homogeneous households, their two costs would be indistinguishable.

every location in both the formal and informal housing sectors:

$$r_{f,k}(d) = \rho \cdot CC + t\Big(\Phi_{f,k} - d\Big), \tag{8}$$

$$r_{i,k}(d) = (\rho + \lambda_k) \cdot CC + t\left(\Phi_{i,k} - d\right). \tag{9}$$

To determine the difference in radii between formal and informal urban development, we can use equations (8) and (9) and the fact that, at any fixed distance from the CBD, rent differences across the two sectors[7] must equal the disutility cost from informal housing, Ψ_k:

$$\phi_k := \Phi_{f,k} - \Phi_{i,k} = \frac{\lambda_k \cdot CC + \Psi_k}{t}. \tag{10}$$

This relation is important: it means that as the city grows, the difference in radius between the two sectors stays constant. The city size and whether a city will actually develop an active informal housing market will depend on the total demand for housing in the city, as well as land availability.

Urban sprawl and the use of space. As in Saiz (2010), we assume that a fraction $1-\Lambda_k$ of land is not available for occupation (formal nor informal) due to geographical restrictions, such as steep slopes or bodies of water. We add a new feature to the model and assume that out of the developable land Λ_k, a share $(1-\Omega_k)$ is rendered dormant in the process of formal urban development – due to bankruptcy, litigation in family successions, general legal disputes, or the presence of public properties without development plans. Although formally vacant, this land is available for informal development.[8]

Given the geometry of the circle, we can tie population to the radius of each sector:

$$\pi \Phi_{f,k}^2 \Lambda_k \Omega_k = H_{f,k} \tag{11}$$

$$\pi \Phi_{i,k}^2 \Lambda_k \left(1 - \Omega_k\right) = H_{i,k}. \tag{12}$$

Noticing from equation (10) that $\Phi_{f,k} = \phi_k + \Phi_{i,k}$ and summing equations (11) and (12), we get a quadratic equation that determines city size as a function of the total population:

[7] From equation (2), $r_{f,k}(d) - r_{i,k}(d) = \Psi_k$.

[8] This is in contrast to other approaches to modeling informal housing (e.g., Brueckner and Selod, 2009), in which legal and illegal housing compete for land.

$$\pi\Lambda_k \left[\Phi_{i,k}^2 + 2\Omega_k \phi_k \Phi_{i,k} + \Omega_k \phi_k^2\right] = H_k, \tag{13}$$

with positive solution:

$$\Phi_{i,k} = \sqrt{\frac{H_k}{\pi\Lambda_k} - \Omega_k (1 - \Omega_k) \phi_k^2} - \Omega_k \phi_k, \tag{14}$$

which determines the informal city radius and, through equation (10), also the formal radius $\Phi_{f,k}$. The condition that ensures a positive solution for $\Phi_{i,k}$ reveals a necessary and sufficient requirement for the presence of slums, which we highlight in Theorem 1.

Theorem 1 *Cities feature some informal housing (slums) if, and only if,*

$$\frac{H_k}{\pi\Lambda_k\Omega_k} \geq \phi_k^2. \tag{15}$$

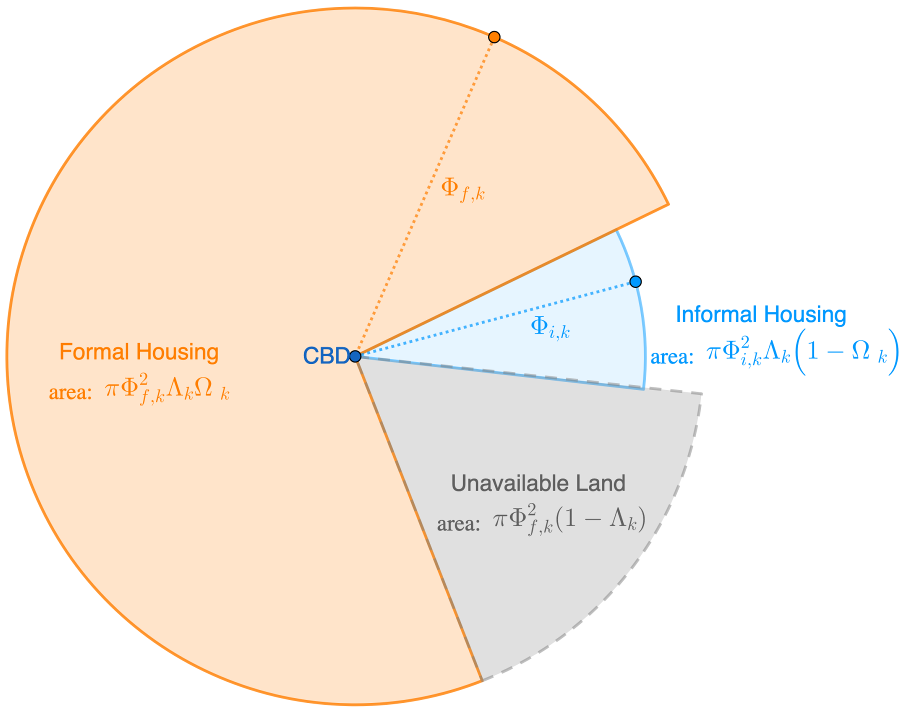
The above condition fails whenever one can actually fit the entire population, H_k, inside a formal city with a radius less than ϕ_k (i.e., in a city with no informal housing). Once that threshold is crossed, the distance to the city's edge is sufficiently large so that informal housing near the city center leads to a commuting time gain that more than compensates the two costs of informal housing (eviction and disutility costs). Moreover, lower eviction rates for informal dwellings (λ_k) imply smaller ϕ_k, meaning a more likely emergence of slums.

There are two directly observed channels for achieving the sufficient urban sprawl that triggers informal land use: population and geographical restrictions. Larger populations (H_k) naturally require more land and make more likely that condition (15) holds. Similarly, tighter geographical restrictions (lower Λ_k), imply the need for more urban sprawl to accommodate a given population, making the condition more likely to hold as well. This relation between city population, geographical restrictions, and the presence of informal housing is a testable prediction of our model we take to the data in Section 3.1.

A similar argument should relate empirically the vacant land share $(1 - \Omega_k)$ and eviction rates (λ_k) to the presence of informal housing, however those variables are not directly observed in the data and only indirect tests are possible.

We summarize the geometry of our circular city model in Figure 1. The diagram displays the formal and informal housing portions of the city, and their different radii, as well as the unavailable land area. We choose to plot each land class area in separate circular sectors for illustrative purposes. The model also accommodates the interpreta-

Figure 1: Circular City with Slums



tion in which each class is evenly spread across the circle according to their respective proportions.

We next investigate the relation between model variables and the informal household ratio, defined as the share of informal households in the total population:

$$IHR_k := \frac{H_{i,k}}{H_k} = \frac{(1-\Omega_k)\Phi_{i,k}^2}{(1-\Omega_k)\Phi_{i,k}^2 + \Omega_k\Phi_{f,k}^2}. \tag{16}$$

Theorem 2 establishes results regarding the presence of informal settlements in cities that are important for our empirical exercise. The proof is in Appendix A.

Theorem 2 *In cities with informal housing, the informal household ratio (IHR_k) is:*

(i) increasing in total population (H_k);

(ii) decreasing in developable land (Λ_k);

(iii) increasing in the vacant land share $(1-\Omega_k)$*, whenever* $\Omega_k > 1/2$*;*

(iv) decreasing in the eviction hazard rate of informal housing (λ_k).

These relations are intuitive. First, since the distance between the formal and informal land occupation radii is constant, as the population grows, a higher share of dwellers live in informal housing. The same intuition applies to the effect of available land (Λ_k). If the city has less available land, it necessarily spreads out more and again, since the distance between formal and informal land occupation radii is constant, this means a higher share of people living in informal housing. Also intuitively, the higher the share of land that is left vacant in the process of urban development, the more spread out is formal development and the more facilitated is the presence of informal housing. Last, a higher eviction rate (λ_k) implies a smaller informal area and informal land occupation.

Empirical model of housing supply. We now discuss the housing supply elasticity in our model. This elasticity includes terms that point to the effect of geographical constraints, present in Saiz (2010), and a novel effect related to informality.

Due to data limitations that we further discuss in the next section, we cannot perfectly identify housing units that are formal or informal. Therefore, we focus on the average behavior in the rental market and on an aggregate notion of housing supply elasticity.

Notice, then, that average rent in city k is given by

$$\overline{r}_k = IHR_k \cdot \overline{r}_{i,k} + (1 - IHR_k)\overline{r}_{f,k}, \tag{17}$$

where $\overline{r}_{i,k}$ and $\overline{r}_{f,k}$ are, respectively, the average rent for informal housing and the average rent for formal housing in city k.

We are interested in the inverse supply elasticity,

$$\beta_k = \frac{d \ln \overline{r}_k}{d \ln H_k}, \tag{18}$$

i.e., the total price response to a shift in the inelastic demand for housing (H_k) across both formal and informal sectors.

In Appendix B we discuss the non-linear expression underlying the supply elasticity in the model. It ultimately depends on all parameters affecting the relative desirability of informal housing, such as the eviction hazard (λ_k), dormant land available for informal use ($1 - \Omega_k$), and the utility penalty from informality (Ψ_k). The economic effect of these parameters is mediated by equilibrium informality, which is endogenous but can be more precisely measured than these quantities.

Rather than focusing on the particular structural relationship implied by the model, we focus on a parsimonious empirical description of the cross-sectional relationship between supply elasticity, land availability, and equilibrium informality. For that, we approximate the supply elasticity as

$$\beta_k = \beta^S + \beta^{LAND}(1 - \Lambda_k) + \beta^{SLUM} IHR_k, \tag{19}$$

which is a combination of an intercept that is shared by all cities, a second term that incorporates the role of land unavailability (as in Saiz, 2010), and a third term that summarizes how informal land use affects supply and prices. We expect $\beta^{LAND} > 0$ and $\beta^{SLUM} < 0$, so that land unavailability increases urban sprawl and prices, while informal land use contains it, reducing the demographic pressure on rents (see Appendix B for more details).

Our empirical inverse housing supply equation is then:

$$\Delta ln(\tilde{r}_k) = \underbrace{\left(\beta^S + \beta^{LAND} \cdot (1 - \Lambda_k) + \beta^{SLUM} \cdot IHR_k\right)}_{\text{Inverse supply elasticity}} \cdot \Delta ln H_k + \gamma_{reg(k)} + \epsilon_k, \tag{20}$$

where Δ means change between 1991 and 2010, $\tilde{r}_k$ is a measure of median rent in the city. We take differences over time to allow for city-specific fixed effects on price levels. We allow for region-specific trends using regional dummies, $\gamma_{reg(k)}$.[9] In our main specification we will focus only on cities with slums, as our goal is understanding the effect of the intensity of informality in the housing supply.[10]

Theorem 2 suggests that in addition to the usual simultaneity bias in the supply equation that relates quantity to the unobserved supply shifter, we have an additional source of endogeneity through IHR_k. That theorem states that IHR_k is directly related to quantity and will therefore also be endogenous in the supply regression. However, Theorem 2 also suggests a solution.

It shows that we can explore exogenous variation in the data that would be correlated with land use, and thus both eviction rates and the share of dormant land, to identify the role played by informality in the housing supply curve. This is the path we explore in our empirical approach.

[9]The official classification assigns Brazilian states to one of 5 regions: South, Southeast, Center-West, Northeast, and North. We use this classification, but group North and Northeast into a single region due to sample size concerns.

[10]For cities without slums, the supply relation (20) would become similar to the specification in Saiz (2010). An investigation of the housing supply in the context of cities without slums would then be a direct application of the existing approach and in theory possible. However, in our application, we would then rely solely on our demographic instrument, which happens to perform poorly for cities without slums, threatening the identification of the supply equation for this particular set of cities.

2 Data and identification

2.1 Data

We define a local real estate market by considering both the strength of economic ties across cities and their physical proximity. We follow the Brazilian Institute of Geography and Statistics (IBGE) official grouping of cities into "Arranjos Populacionais e Concentrações Urbanas," which is a close equivalent to the Metropolitian Statitical Area (MSA) concept in the United States.

A particular municipality is grouped with one or more neighbors into an *arranjo* whenever one of the following two conditions is satisfied: (i) a daily flow of workers and students that is either above 10,000 people or above 17% of the city's population; or (ii) urbanized areas that lie less than three kilometers apart. Our sample consists of all *arranjos* with more than 100,000 residents in the 2010 National Population Census. We refer to these units as MSAs. We collect demographic data and the household count across the 1991 and 2010 censuses. Only occupied houses are included in this dataset. We also collect housing rents from an extended census survey administered to a 25%-coverage sample of census households.

Our analysis is based on rental data. We use it as a proxy for house prices whenever necessary, in an approach that can be justified by a present-value house pricing model. In order to mitigate concerns related to composition effects related to housing quality, we construct measures of city rents that control for observable quality measures as we detail next.

First, we normalize reported rents by the total number of rooms (not only bedrooms) in a house. We also exclude single-room units from the analysis.[11] Figure 2a displays the distribution of the number of rooms in a house for 1991 and 2010. This distribution is stable over this time frame, mitigating concerns for composition effects in the rent-per-room measure. Figure 2b displays the average rent-per-room as a function of the number of rooms. It shows a mild U-shaped pattern. Notice that most of the distribution from the Figure 2a is concentrated around four and five room units. In this range, rent-per-room does not vary strongly with the underlying number of rooms.

In a second step, we further control for other housing quality measures that are priced in rents and may vary as cities grow and become more informal. We create rent indices based on an hedonic model for rent-per-room as, e.g., Baum-Snow and Han (2021). We

[11] We view these units as indicative of a single bedroom rental in a multi-room housing unit. They do not include bathrooms or kitchens and their rent-per-room is significantly higher than comparable multiple-room units. They are approximately 1% of the original sample.

regress log of rent-per-room on a series of dwelling characteristics such as access to electricity, water and sewage services, wall material, and multi-family occupancy of buildings. Importantly, the dwelling characteristics dataset used here contains information on rents, number of rooms, and some characteristics of the building, but not if the dwelling in question is located in a slum. Estimates from the hedonic model are displayed in Table C.1. We use the residuals of this regression as our measure of quality adjusted rent-per-room. Finally, we take the difference of the median MSA quality adjusted rent-per-room between 2010 and 1991 to construct the dependent variable in our housing supply equation (20).

Due to the informal nature and precarious conditions that essentially define slums, data on these settlements are often rare and inaccurate. Nonetheless, IBGE provides an official slum definition using some characteristics of informality, public utilities, and urban development. Its technical name is a *subnormal agglomeration*. Most of these agglomerations do not have access to essential public services and the dwellings therein are generally arranged in a disorderly and dense manner.

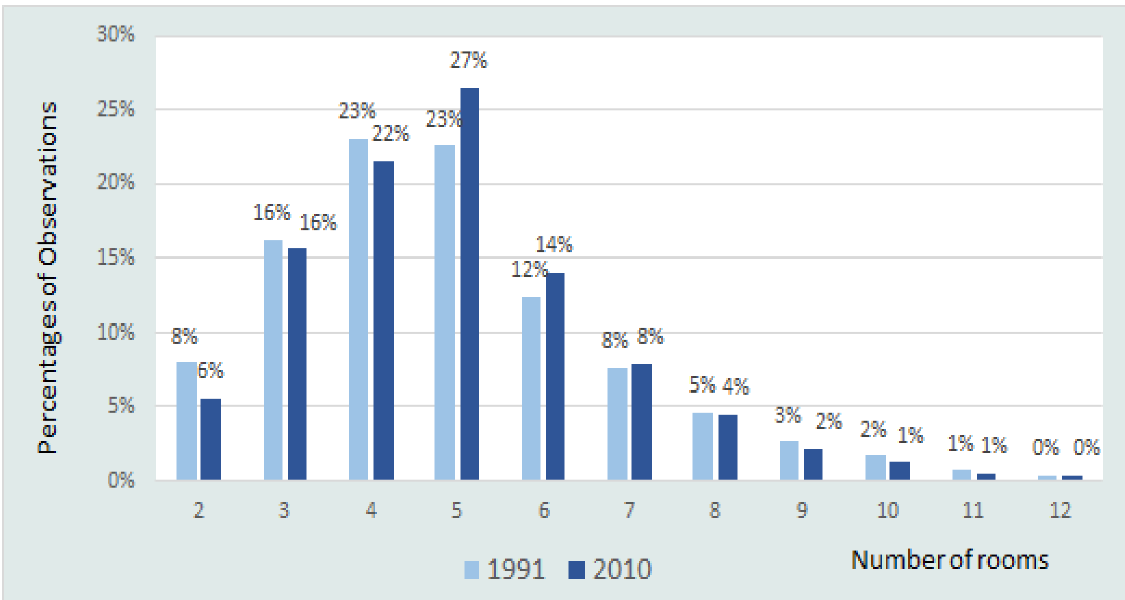
The identification of a subnormal agglomeration is based on the following criteria: (i) illegal occupation of underlying land; *i.e.*, unauthorized construction on land formally owned by (public or private) third-parties at the present moment or in a recent period (this nests the case of a property title granted within the previous ten years); and (ii) at least one of the following two characteristics: (a) urban development outside current standards – narrow and irregularly aligned roadways, plots of unequal sizes and shapes, and constructions not validated by public authorities; or (b) precariousness in essential public services.

IBGE has made an effort to improve the quality of their measurements regarding the criteria above in the 2010 census.[12] The downside of this improvement is that classifications are not comparable across censuses. Therefore, We use throughout this paper the most recent and reliable data from the 2010 census to measure IHR_k. Our interest here is exploring the role played by informality in the supply curve and not on the dynamics of informality in each city, so we believe this is not a severe limitation to our approach. Moreover, it is not possible to directly relate the subnormal classification to the household-level dataset containing the rental variable we use to construct prices.

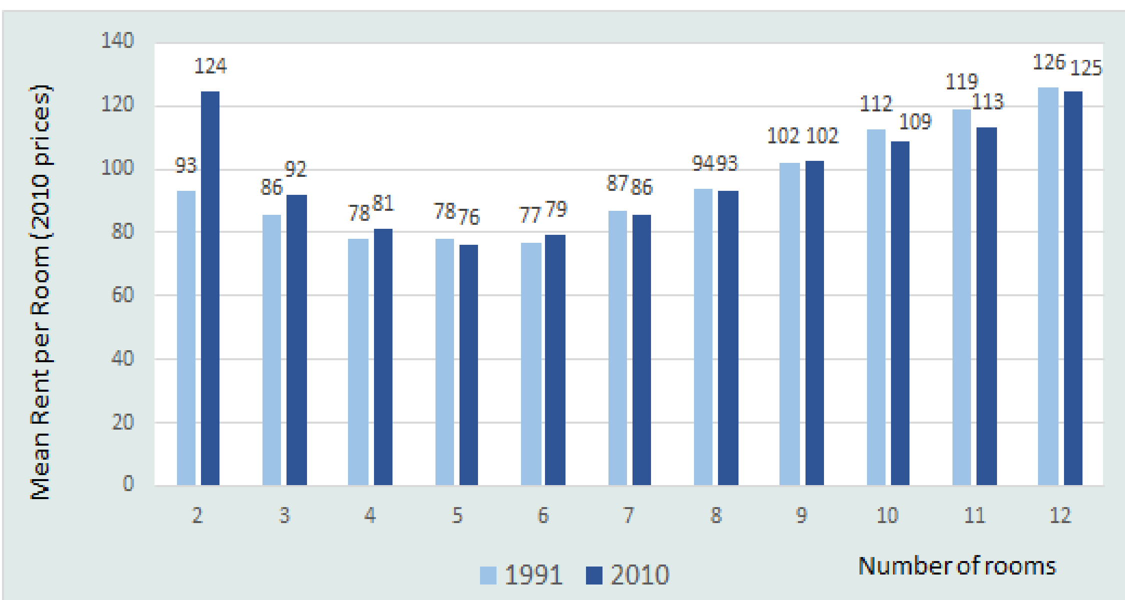
[12]In 2010, IBGE started using high-resolution GIS data from satellite imagery to identify census blocks and non-conforming urban development. Moreover, information on land ownership from municipal records were used in large scale for the first time. They have further refined the initial mapping through Municipal Commissions of Geography and Statistics in 350 cities to add, remove, and define these subnormal agglomerations. Field technicians have also evaluated and verified the slum characteristics as defined above. For a discussion, see IBGE (2011) and Mation et al. (2014).

Figure 2: Number of Rooms

(a) Distribution by Number of Rooms



(b) Mean Rent per Room



This data limitation prevents a separate analysis of formal and informal housing supplies without strong assumptions.

We complement that data with a dataset on geographic constraints on construction that we develop for this study. To construct this dataset, we use geographic information system (GIS) data for Brazilian MSAs and their vicinities. We compute inclination measures using the *Instituto Nacional de Pesquisas Espaciais* satellite-based data of the digital elevation model, that provides slope maps at a 30-meter resolution. We additionally use the U.S. Geological Survey satellite-based geographic data to characterize bodies of water at a 0.5-kilometer resolution.

Using inclination and water coverage data, we provide criteria for the unavailability of land for development in a given location. Our main measure defines the share of unavailable land in a city as the fraction of map pixels within a 10-km radius around the city center that are either covered by water (lakes, rivers, and sea) or too steep to allow construction without significant hurdles. Our choice of a 10-km radius avoids large overlaps between circles defined around cities without significant economic ties.

Table 1: Share of Census Blocks Below Given Slope - 2010 Census (%)

	Brazil		Northeast and North		South		Southeast	
Slopes	Non-slums	Slums	Non-slums	Slums	Non-slums	Slums	Non-slums	Slums
< 15%	83	70	92	82	88	75	77	61
< 20%	92	82	98	93	93	84	89	74
< 30%	98	94	99	99	98	94	97	89
N	132,361	15,539	26,405	5,703	18,518	854	76,591	8,709
Type Share	89%	11%	82%	18%	96%	4%	90%	10%

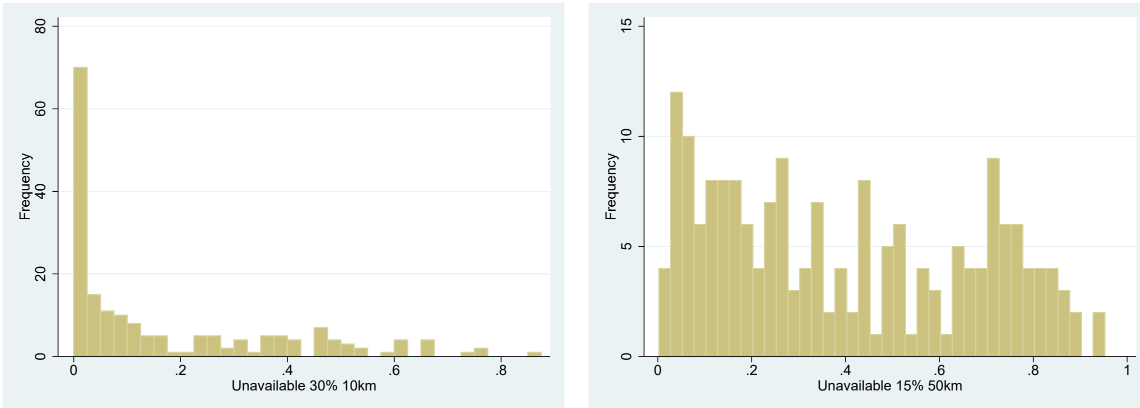
Note: Percentages in the middle of the table are the share of census blocks with slopes below 15%, 20%, and 30%. Type Share is the ratio of non-slums (or slums) blocks in the sample. Only census blocks of MSAs with slums are included in this descriptive analysis.

For our main measure, a map pixel is deemed too steep whenever it features an inclination above 30%.[13] As Table 1 shows, only 6% of the census blocks (*setores censitários*) in our sample have both a subnormal agglomeration and inclination above 30%, while 98% of census blocks without slums have slopes below 30%.[14] Additionally, the 30% slope cutoff is meant to be conservative, only deeming unavailable land where construction

[13]The slope of a pixel in the map is defined as the elevation gain over that pixel divided by its length.

[14]Moreover, federal law number 6766 regarding "Parcelamento do Solo Urbano" (Urban Land Development Plan) from 1979 forbids construction above that threshold.

Figure 3: Distribution of Unavailable Land Across Cities



(a) Unavailable Land: 30% slope and 10-km radius definition

(b) Unavailable Land: 15% slope and 50-km radius definition

Note: Frequency of MSAs by fraction of land unavailable in MSA center's vicinity. The fraction of unavailable land used in (a) is the share of pixels inside a circle of 10-km radius which is either covered by water or with slope higher than 30%. For (b), we consider a 50-km radius circle and 15% steepness as a threshold.

would be very challenging, even using technologies that are likely to be too costly for an informal settlement. To evaluate the robustness of our results, we also provide an alternative measure following Saiz (2010), using a 50-km radius and a 15% slope as a cutoff. Figure 3 shows the distribution of the share of unavailable land across cities for these two alternative measures.

In Table 2, we report a summary of descriptive statistics for the main variables in our supply equation. We separate those statistics for MSAs with slums and for the extended sample of all MSAs. We first note the large increase in population in the MSAs over this period. Variation in population is key to identify the supply curve in our empirical specification. In line with our model, MSAs with slums are bigger and feature higher rents than MSAs without slums, a stylized fact we formally document in Section 3.1. Moreover, MSAs with slums present more unavailable land, which is due primarily to differences in the area occupied by water bodies rather than differences in the presence of steep terrain.

2.2 Identification by instrumental variables

The identification of housing supply relation in equation (20) needs to overcome the usual problem of demand and supply simultaneity, which makes prices and quantities endogenous. Here we follow the classic approach in supply estimation and use an instrumental

Table 2: MSAs Descriptive Statistics

	MSAs with Slums		All MSAs	
	mean	sd	mean	sd
Households 1991 (thousands)	193.2	507.0	111.1	364.0
Households 2010 (thousands)	329.3	777.3	190.4	561.1
Rent per Room 1991 (MSA median)	58.5	16.5	54.0	17.5
Rent per Room 2010 (MSA median)	68.8	16.7	64.1	16.4
% Slum Households 2010 (IHR)	8.2%	8.5%	4.0%	7.2%
Unavailable 15% 50km	47.9%	27.1%	39.5%	27.3%
Unavailable 30% 10km	25.0%	24.0%	17.1%	21.2%
Water 10km	15.9%	21.9%	9.1%	17.8%
Water 50km	18.0%	22.1%	10.8%	18.2%
Slope 15% 50km	29.9%	25.8%	28.6%	25.0%
Slope 30% 10km	9.1%	16.0%	8.0%	14.1%
N	91		185	

Note: Rent per room measured in 2010 Brazilian reals (BRL).

variable arguably unrelated to supply-side unobservables (exclusion restriction), but that is a demand shifter and therefore correlated with changes in the number of households (relevance condition). Since our housing supply equation is estimated in differences between 1991 and 2010, our instrument needs only to be uncorrelated with changes over time in supply-side unobservables at the city level.

We are faced with yet another identification challenge, as the informal household ratio (IHR_k) is also an endogenous variable. Unobservable variables related to supply shocks may shift IHR_k in different cities. To circumvent these problems, we also adopt an instrumental variables approach. In the particular case of IHR_k endogeneity, we follow the guidance of the model from Section 1, which indicates that exogenous variations in either eviction risk or in the share of dormant inner-city land influence the extent of informal housing.

We instrument the housing quantity and IHR_k variables with demographic demand shifters and total area under federal ownership in the cities. This ensures proper identification of housing supply under the assumptions that: (i) demographic composition affects demand for housing through expected population growth, for instance, but does not affect housing supply directly; and (ii) federal land ownership is essentially fixed, for historical reasons, and does not respond to unobservable shocks that affect the housing supply. To mitigate concerns about endogeneity, we add regional fixed effects in the empirical specifications that follow.

Demographic instruments. We propose here two demand shifters that are anchored on the demographic composition of our cities in 1991 and on national mortality rates.

Demography is a key driver of city growth. In the case of Brazil, where incomes are low and transportation costs significant, this is heightened by a lower migration rate. In itself, a low mobility does not imply a failure of a frictionless spatial equilibrium across cities, but it may be a sign that the equalizing force of migration is not present in its full strength. Chauvin et al. (2017) supports the finding that mobility has been traditionally lower in Brazil compared to the US, but not low enough, in their view, to compromise the applicability of the Rosen-Roback spatial equilibrium model.

Considering this, we build instruments based on an estimate of the rate of natural increase (RNI) for each MSA that excludes in and out-migration and contemporaneous fertility responses. This is important because migration and fertility may respond to innovations in housing supply unobservables through traditional demand and supply simultaneity. We construct that rate restricting attention to the population aged 20 or older, who are likely to demand housing. For each MSA, we compute:

$$\Delta lnPD_k = \ln\left(\sum_{a=0}^{70}(1-h(a))Pop_k^{1991}(a)\right) - \ln\left(\sum_{a=21}^{90} Pop_k^{1991}(a)\right), \tag{21}$$

where $Pop_k^{1991}(a)$ is the total population of age a in MSA k in 1991 and $h(a)$ is the predicted mortality rate between 1991 and 2010 for people who are a years old in 1991. The first term in the difference is the predicted population above 20 years old in 2010, which are all those alive in 1991 adjusted by predicted mortality for each age cohort. The second term is just the population above 20 years old in 1991. The log difference between the two terms is thus the RNI.

This predicted (by demography) increase in population, $\Delta lnPD_k$, should be correlated with changes in the demand for housing between 1991 and 2010 and thus a good candidate to instrument changes in quantity in our supply equation. The exclusion restriction will be satisfied if $\Delta lnPD_k$ is uncorrelated with innovations in housing supply unobservables between 1991 and 2010. This is plausible given that the demographic composition in 1991 is due to factors – e.g., fertility and migration – anchored at least in the few decades before 1991 and should be uncorrelated with innovations in supply-side unobservables between 1991 and 2010, such as changes in construction costs and land use regulation. Moreover, in order to mitigate concerns of local endogenous mortality responses to unobservable factors affecting housing supply, we use the national level mortality rates.[15]

In addition to $\Delta lnPD_k$, we use also include as instruments the level of the predicted number of households in 2010, PD_k^{2010}. This level is correlated, as predicted by our model, with IHR_k, but arguably unrelated to changes in the housing supply between 1991 and 2010. The argument for the exclusion restriction is the same as the one laid before since PD_k^{2010} also only depends on demographic composition in 1991 and mortality rates between 1991 and 2010. We assess the instruments' first stage in Section 3.2.

Public land instrument. New slums (*favelas*) have been identified in all major cities across the country over the last century. The first *favelas* appeared in the late 19th century, in downtown Rio de Janeiro. Along with industrialization and a process of urbanization that picked up in the 1940s, these unauthorized settlements became commonplace in major Brazilian cities. Earlier policies based on attempts to reclaim occupied land and resettle *favela* residents in public housing projects have been mostly abandoned, due to a combination of fiscal constraints and an international policy shift away from public housing provision. Nowadays, local governments focus more on transfers of property rights, infrastructure development, and the provision of essential service improvements without resettlement.

The longstanding presence of slums has been suggested as a sign of state tolerance and ineffective regulation in metropolitan areas. For these reasons, public land holdings are also widely believed to be more susceptible to invasion. According to Flood (2006), "49% of land invasions in Sub-Saharan Africa, 60% in North Africa, East Asia, and West Asia and 90% in South Asia occur on public land" (Shah, 2014), while according to Brueckner and Selod (2009), only 40% of "land invasions" in Latin America and the Caribbean occur on private land.

In the case of Brazil, federal land holdings were defined in the constitutions of 1831 and 1946 and reaffirmed in the current 1988 Constitution.[16] Regularization of low-income housing built on public properties has been one of the goals of the federal govern-

[15] We also explored a version of these demographic instruments using state and regional level mortality rates. Our main estimates barely changed under this alternative demographic instruments.

[16] See decree-law number 9,760, of 1946, and article 20 in the 1988 Constitution. Typically, land that neighbors the ocean, lakes, and rivers has been attributed to the Federal Union. Publicly available government budgets show negligible amounts raised from land sales and there is a widespread recognition of the hurdles and limitations to any attempted sale. Only recently, starting in 2019, an effort to identify and catalogue federal real estate began. The dataset of federal land holdings we use was actually built as part of that effort.

ment, which initiated processes of ownership transfer to approximately 500,000 households between 2003 and 2010 (Secretaria do Patrimônio da União, 2010).

Given the higher vulnerability of public land to invasions, we use the total share of federal land holdings in each MSA as our main instrument for informal housing ratio.[17] We do not include land held by municipalities and states, as these variables are likely to be subject to the influence of endogenous urban development policies and other shocks that influence local real estate markets.[18]

First-stage specifications. Naturally, we have one first stage for each of the three endogenous regressors in the supply equation (20), that is, ΔlnH_k, $(1-\Lambda_k)\cdot\Delta lnH_k$, and $IHR_k\cdot\Delta lnH_k$. We summarize those first-stage relations and our use of the instrumental variables discussed above in the next equation:

$$Y_k = \alpha_1\Delta lnPD_k + \alpha_2(1-\Lambda_k)\cdot\Delta lnPD_k + \alpha_3 PP_k\cdot\Delta lnPD_k + \alpha_4 PD_k^{10}\cdot\Delta lnPD_k + \alpha_{reg(k)} + \epsilon_k, \tag{22}$$

where $\Delta lnPD_k$ is the demographic growth projection, PD_k^{10} is the projected number of households for 2010, PP_k is the the ratio of public property to the MSA's area, and $\alpha_{reg(k)}$ are regional fixed effects. The dependent variable Y_k is one of three endogenous regressors in equation (20).

3 Results

In this section, we report our results, starting with an analysis of the determinants of the presence of informal housing, followed by a discussion of our two-step housing supply estimation and elasticities.

3.1 The extensive margin: Which cities have slums?

We first investigate the empirical implication of our model with respect to the presence of slums in cities. Theorem 1 establishes that both population and geographical restrictions are related to urban sprawl and thus to the presence of informal land use. This allows a direct test of the model. Additionally, an indirect test of the model and of our use of the public land instrument is possible considering the relation between this instrument and

[17]Although the share of federal land holdings is historically determined, we found no evidence of correlation between this measure of public properties and city age or city size, which reinforces the exogeneity argument for this variable as an instrument in a housing supply equation.

[18]Article 185 in the Constitution of 1988 states that urban policy will be approved and implemented by municipal authorities.

the presence of slums. Federal land ownership is expected to increase the amount of dormant land near the city center and to decrease eviction rates. All else equal, both of these forces make a city more likely to satisfy condition (15), which leads to the emergence of informality.

In Table 3, we report estimates from probit binary response models in which the dependent variable is equal to 1 whenever the city has any slums and is 0 otherwise. The estimates overall support the model implications in a variety of specifications. In columns (1) and (2), we show that land unavailability is correlated with the presence of informal housing as our model indicates. This relation is robust even if we separate the land unavailability effect into its two sources (column 2). We next add population and the public land instrument to the specification. Again, our estimates in column (3) are in line with the model's predictions, as both population and the public land instrument are positively correlated with slum occurrence. However, one could be concerned about the possible endogeneity of the population in this response model.

For instance, people could leave or join a city due to the presence of slums. Therefore, in the specification displayed in column (4), we use the demographically predicted population as a regressor instead, as this variable is at least anchored in past city choice decisions. The estimates are still supportive of the model implications, with small changes in coefficients.

3.2 Housing supply first-stage

We next present direct evidence of instrument relevance in the first stage of our two-stage least squares estimation of the housing supply. We focus on our preferred sample of MSAs with slums. In Table 4, we present the relationship between the number of households in an MSA and our exogenous demographic instrument. Our focus lies on column (2) and the coefficient on the demographic instrument, after controlling for regional dummies. We find a strong relation between the demographically predicted variation in the number of households and the actual change in the number of households. A one percentage point increase in the predicted population growth between 1991 and 2010 increases actual population growth on average by approximately 0.733 percentage point.

In Table 5 we display results from regressions that serve a dual purpose. They aim to attest the relevance of the instruments for IHR_k, but they also serve as an empirical validation of our model. First, we learn that both instruments for the share of informal households are positive and significantly related to IHR_k, after controlling for regional dummies and land unavailability. Point estimates imply that, all else held equal, a 10

Table 3: Which Cities have Informal Housing Clusters?

Probit binary response *Slum Presence Dummy*	(1)	(2)	(3)	(4)
Unav. Land 30% 10 km	2.546***		3.184***	3.555***
	(0.511)		(0.764)	(0.789)
Water 10 km		3.748***		
		(0.814)		
Slope 30% 10 km		1.170*		
		(0.695)		
Log(# Households 2010)			1.309***	
			(0.200)	
Log(Predicted Households 2010)				1.378***
				(0.205)
% Public Land†			0.643*	0.689*
			(0.365)	(0.372)

Note: Standard errors are in parentheses. All specifications use the entire sample of 185 MSAs and include regional dummies for South, Southeast, Center-West, and North/Northeast. † Area of federal public properties without building divided by the MSA urban area. Predicted Households 2010 is the projection for 2010 population considering mortality rates and demographic composition of each MSA in 1991. *, **, and *** indicates significance at the 10%,5%, and 1% levels, respectively.

Table 4: Housing Quantity Instruments

Δ *Households* (ΔlnH_k)	(1)	(2)
Δ *Predicted households* ($\Delta lnPD_k$)	0.661***	0.733***
	(0.206)	(0.253)
Regional Dummies	No	Yes
R^2	0.149	0.917

Note: Robust standard errors are in parentheses. Sample of MSAs with slums (91 MSAs). The Demographic Instrument estimate the RNI of population over 20yo between 1991 and 2010 using mortality rates and demographic composition in 1991. Regional Dummies are South, Southeast, Center-West and North/Northeast. *, **, and *** indicates significance at the 10%,5%, and 1% levels, respectively.

Table 5: Slums Instruments

% Slums/Households (IHR)	(1)	(2)
% Public Land†	0.017**	0.014*
	(0.006)	(0.008)
Predicted Households 2010	0.016**	0.016**
	(0.006)	(0.007)
Unav.30% 10 km	0.164***	0.156***
	(0.037)	(0.036)
Regional Dummies	No	Yes
R^2	0.327	0.694

Note: Robust standard errors are in parentheses. Sample of MSAs with slums (91 MSAs). Regional Dummies are South, Southeast, Center-West and North/Northeast. † Area of federal public properties without building divided by the MSA urban area. Predicted Households 2010 is the projection for 2010 population based on mortality rates and population age distribution of each MSA in 1991. *, **, and *** indicates significance at the 10%,5%, and 1% levels, respectively.

percentage-point increase in the share of land that is federally held leads to an expected increase of 0.14 percentage point in the fraction of households that live in slums, statistically significant at a 10% confidence level. Our other instrument, the predicted number of households, is also positively correlated to the share of informal households at a standard 5% significance level. Second, those regressions also serve as a confirmation of the model predictions set in Theorem 2. They suggest that cities that are large, more geographically restricted, and with more vacant land (measured through public land) do feature higher shares of informal housing, as predicted by our model.

In Tables 4 and 5, we report results that confirm the instruments' relevance in our setting. Notice, however, that these simplified specifications do not represent the formal first-stage equations (22), that feature hard-to-interpret coefficients, due to the presence of interactions between dependent and independent variables. We report the full first-stage equations in Appendix Table C.2.

Instruments are correlated with endogenous variables considering individual t-tests, but excluded instruments F-statistics' for each endogenous variable range between 4.82 and 37.17, which is sometimes below commonly used rules-of-thumb thresholds for weak instruments detection. However, those F-statistics are not sufficient to detect the presence of weak instruments in our setting. Since we have 3 endogenous regressors, the alternative here would be to perform a test using the Cragg-Donald statistic using critical values such as computed by Stock et al. (2005). Those critical values are not computed for our specific case (3 endogenous variables and 4 instruments) and only apply for the homoskedastic case, when in fact the Breusch–Pagan test rejects homoskedasticity at 5% in all first-stage specifications. As far as we are aware, the literature is silent with respect to weak IV testing procedures applicable to our specific setting (Andrews et al., 2019). Our approach is then to proceed with caution and provide inference and estimation results for our supply equation that are robust to the presence of weak instruments. In any case, we are still able to test for underidentification in the next section, because this test is valid under more general conditions.

We also explored longer specifications for the first stage, using more flexible interactions for the exogenous variables. The final results changed only marginally, so we opted for the more parsimonious specification for the first-stage described in equation (22).

3.3 Housing supply estimation results

We present estimates for the inverse supply equation (20) in Table 6, the second-stage of our two-stage least squares estimation. We focus on cities with at least one slum, highlighting the intensive margin of informal housing, as indicated by the model in Section 1.

First, the ordinary least squares results do not indicate any meaningful relation between variables, indicating the potential confounding effects of geography and endogenous informality in housing markets. However, once we use our set of instruments in the specifications for columns (4) to (6), we can interpret the estimated relations as inverse supply equations.

The results from our main specification, which includes regional dummies, are presented in column (6). The estimates indicate that an increase in the population share living in slums leads to a decrease in the inverse supply elasticity that is both economically and statistically significant. Moreover, a higher share of unavailable land increases the inverse supply elasticity. In other words, cities with higher shares of slums and less unavailable land experience a higher supply elasticity (lower inverse supply elasticity) in the housing market. This finding is consistent with Alves (2021), who finds a lower inverse supply elasticity for unserviced houses in Brazil. Excluding regional dummies (column 5) does not lead to any meaningful changes to the interactions of housing supply inverse elasticity with land availability and informality.

We next also consider a specification reminiscent of the one in Saiz (2010), in which we include only the geographic constraint variable and not the IHR. We display this specification in column (4).[19] We find very imprecise estimates for the base elasticity and the geographic constraints effect. The intuition is that the prevalence of slums generates an effect that goes in the opposite direction of the geographic constraints effect and leads to its attenuation. Since geographical constraints and slums are correlated, this yields ambiguous results if these two variables are not properly accounted for. Once we include slums in the specification, the geographic constraint coefficient sign becomes distinguishable from zero.

We test for underidentification in all specifications. The Kleibergen-Paap test can reject the null of underidentification at 5% for our main specification. However this test is not sufficient to discard the possibility of weak instruments. We therefore report an inference result which is robust to the presence of weak instruments. We report p-values for the Anderson-Rubin test of the null that all coefficients on endogenous regressors are zero. We can reject the null at 1% in our main specification. Additionally, in order to mitigate further concerns of instrument validity, we compute Sargan-Hansen tests of over identification and fail to reject the null of valid instruments in columns (5) and (6).

3.4 Robustness

Sample of MSAs. We consider two variations in our sample of cities. First, we consider the sample of all MSAs and not just the ones with at least one slum. We report housing supply estimates for all MSAs in Appendix Table C.5.[20] Although there is an attenuation in the inverse elasticity relationship with land unavailability and informality, the signs of the relations are preserved.

We treat the share of unavailable land as exogenous and one may worry that this measure is correlated with other city features. For instance, coastal cities have more land unavailable but may also feature ports and other specific industrial land uses that may compete with housing. Therefore, we consider a set of specifications in which we focus only on non-coastal cities.[21] We display results for this specification in Appendix Table C.6. The restriction to non-coastal MSAs implies a 34% drop in our sample size, however we still find a significant negative effect of IHR and a significant positive effect of unavailable land.

[19]This specification is run with a shorter list of instrumental variables. We do not include the IHR and use as instruments only the predicted population growth and its interaction with geographical constraints. Therefore, in this case, we end up with a just-identified IV econometric model.

[20]In Appendix Tables C.3 and C.4, we provide regression tables in support of instrument relevance for, respectively, the change in the number of households and IHR in the sample of all MSAs. Estimates go in the same direction as those in our baseline sample of MSAs with slums. Some instruments seem even stronger in the sample of all MSAs, *e.g.*, the public land instrument.

[21]We thank an anonymous referee for making this suggestion.

Table 6: Supply Curve 1991-2010: MSAs with at least one slum

Δ*Rent index* ($\Delta ln(\tilde{r}_k)$)	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)
ΔlnH_k	−0.008	−0.083	0.001	0.393	−0.008	0.457
	(0.138)	(0.151)	(0.139)	(0.589)	(0.333)	(0.490)
$Unav.30\%10km \cdot \Delta lnH_k$	0.149	0.270*	0.175	0.150	0.806***	0.825**
	(0.131)	(0.149)	(0.154)	(0.178)	(0.260)	(0.336)
$IHR_k \cdot \Delta lnH_k$		−0.377	−0.132		−3.428***	−3.813**
		(0.392)	(0.355)		(1.170)	(1.554)
Regional Dummies	Yes	No	Yes	Yes	No	Yes
AR-p				0.3506	0.0000	0.0004
KP-p				0.0035	0.0869	0.0315
Overid-p				Just-ID	0.5195	0.2414

Note: Heteroskedasticity robust standard errors in parentheses. This table displays estimates of the supply equation (20) for the sample of MSAs with slums (91 MSAs). The notation Δ before the variable means change between 1991 and 2010. The dependent variable is change in median rent-per-room index in all specifications. ΔlnH_k is change in the log of the number of households. IHR_k is the percentage of informal houses in 2010. $Unav.30\%10km$ is the share of unavailable land ($1-\Lambda_k$, in eq. (20)), in which we use a 10-km radius from the CBD and consider all land above 30% inclination as unavailable. AR-p indicates p-values for the Anderson-Rubin test of the null that all coefficients on endogenous variables are zero. KP-p indicates p-values for the Kleibergen-Paap test of the null of underidentification. Overid-p indicates the p-value for the Sargan-Hansen test of the null that instruments are valid. *, **, and *** indicates significance at the 10%,5%, and 1% levels, respectively.

Geographical constraints. The definition of unavailable land is somewhat arbitrary. Although considering water bodies as unavailable is uncontroversial, the choice of radius from the CBD to consider in the computation and the slope above which land is considered unavailable may be not. We choose a radius (10km) and a slope (30%) that we believe are more appropriate for our context than, for instance, the choice in Saiz (2010) (50km radius and 15% slope). A 50-km radius is probably too large for most Brazilian MSAs, which are notorious for their bad transportation infrastructure. Moreover a 15% slope is not very restrictive either, as many urban dwellings are built on slopes above 15% (see Table 1).

As a robustness check, we display in Appendix Table C.7 results for our housing supply estimates using a less tight measure of unavailable land based on a 50km radius and 15% slope. In comparison with our main specifications, the signs of the coefficients remain the same, but the effect of unavailable land is attenuated when we consider the looser definition of geographical constraint.

Estimation by LIML. As a final robustness check, we estimate the supply equation by LIML instead of 2SLS. The LIML estimator will be less biased than 2SLS when instruments are weak (Hansen, 2022, p. 393). Moreover, in over-identified models, LIML inference will be less sensitive to the presence of weak instruments (Stock et al., 2005). We display housing supply estimates by LIML in Appendix Table C.8. Coefficient estimates are in the same direction and support the same qualitative conclusions as our leading specification.

3.5 Elasticity estimates

We compute inverse housing supply elasticities estimates at the MSA level using our leading specification, as reported in Table 6, column (6). Given equation (20), for each MSA k, our estimated inverse housing supply elasticity is then:

$$\hat{\xi}_k^S = \hat{\beta}^S + \hat{\beta}^{LAND} \cdot (1 - \Lambda_k) + \hat{\beta}^{SLUM} \cdot IHR_k. \tag{23}$$

In order to understand the relative importance of each component of the supply elasticity in the production of the total elasticity, we decompose the variance of MSAs elasticity and present the results in Table 7. Geographic constraints to city expansion previously considered in the literature (Saiz, 2010) respond only for 49% of the total variance in inverse supply elasticities. Meanwhile, ignoring geography and considering only the prevalence of slums actually increases the variance by 33%. This is because geographic constraints and slums are positively correlated, but are opposing forces in the production of inverse elasticities. This highlights the importance of considering both the prevalence of slums and geographic constraints to explain the range of price responses to quantities along the housing supply curve.

We present a full list of inverse housing supply elasticities for all MSAs with at least one slum in Appendix Table D.1. We note a vast range of inverse elasticities, which highlights the dispersion in price responses that are expected given housing demand shocks. For 7 out of 91 MSAs, we find negative inverse elasticity point estimates. Those are MSAs that featured a high share of housing informality and/or very low shares of unavailable land, under which our linear empirical specification for the inverse elasticity produces implausible negative estimates. However inverse elasticity point estimates for six of those MSAs are less than one standard error away from zero, which implies that we cannot reject a small positive inverse elasticity for those cities.[22]

Average point estimates for inverse elasticities are 0.35. If we only include MSAs with positive point estimates the average is a 0.4. For MSAs with positive point estimates, the elasticity (not inverse) estimates are, on average, 4.22, with a high 6.67 standard deviation, which is mainly driven by a few MSAs that feature close to zero inverse elasticities.

We therefore find higher elasticities on average than Saiz (2010), who finds a 2.5 average supply elasticity. This difference between supply elasticities in American and Brazilian MSAs can be hardly explained by differences in geographical features in the two countries, as American and Brazilian cities feature on average similar shares of unavailable land.[23] This difference however can be rationalized by the widespread presence of informality in Brazil compared to the US. Informality as we have documented tends to increase (decrease) the (inverse) housing supply elasticity.

According to our estimates, a city with the average inverse supply elasticity of 0.40 would face an inverse elasticity of 0.13 if it had a one standard deviation (8.5 p.p.) higher IHR. Inverse elasticities and the effect of informality here are in a range comparable to Alves (2021)'s estimates – specifically, a 0.4 inverse elasticity for formal housing and 0.07 for informal housing in Brazil – which he estimates using a different definition of housing informality and empirical approach.

We illustrate the importance of informal housing with the case of two MSAs from

[22]The exception is Belém MSA, with a inverse elasticity point estimate of -1.22 (0.71 standard error). Belém is a clear outlier in our sample with an IHR of 54%, which is more than fives standard deviations away from the average IHR. This extreme informality ends up pushing the inverse elasticity to an unreasonable negative value.

[23]In Brazil (with our preferred specification) 25.0% of land is unavailable on average and in the U.S., 26.4% (from Table 1 in Saiz, 2010). Histograms of unavailable land were also similar across the two countries.

Table 7: Inverse Elasticity Variance Decomposition

	Total (slums + geography) $\hat{\beta}^{S}+\hat{\beta}^{LAND}\cdot(1-\Lambda_k)+\hat{\beta}^{SLUM}\cdot IHR_k$	Slums only $\hat{\beta}^{S}+\hat{\beta}^{SLUM}\cdot IHR_k$	Geography only $\hat{\beta}^{S}+\hat{\beta}^{LAND}\cdot(1-\Lambda_k)$
Variance	0.079	0.105	0.039
(%)Total variance	100%	133%	49%

Note: Variance decomposition of inverse supply elasticity. We use estimates displayed in Table 6 column (6). Formulae for inverse supply elasticity (based on eq. (23)) are displayed in the respective columns.

the Brazilian coast, with severe geographical constraints but a totally different housing supply: Florianopolis and Baixada Santista. Florianopolis' center lies on a mountainous island. Sixty-one percent of the land within a 10-km radius of the city center is unavailable, but only 3% of Florianopolis' households live in informal housing. In the absence of a substantial informal housing market, topography alone produces a high inverse elasticity of 0.86, the second highest in our sample. Meanwhile, Baixada Santista is squeezed between the sea and the mountains, and 55% of its area is unavailable due to geographic constraints. However, it has an informal market share of 18%, which reduces its inverse elasticity to 0.23, showing the strength of the informality effect for the supply elasticity estimates.

In another interesting contrast, Rio de Janeiro and Sao Paulo, the two biggest cities in Brazil, have a similar informal housing share in their metropolitan areas, but a very different unavailable land ratio within a 10-km radius (58% and 0%, respectively). In this case, geographic constraints have a remarkable impact on Rio's supply curve, raising its inverse price elasticity to 0.42, while in Sao Paulo the inverse elasticity is 0.06, one of the lowest in our sample.

We hope the approach proposed here and these supply elasticities estimates may help economists and policy makers in several applications related to housing markets in developing countries. In this spirit, as an illustrative application, we show in Appendix D how we can use our elasticity estimates to simulate predicted trajectories implied by natural population growth. In Appendix Table D.1, we display the predicted natural population growth and implied housing price increases for all cities with at least one slum. The average predicted change in the number of households is 32%, while the average predicted price increase is 12%.

4 Conclusion

In this paper, we analyze housing supply in Brazil, with its formal and informal sectors, a prevalent duality in many developing countries. We extend a monocentric city model to incorporate this dual housing market. This model guides our empirical application and we find that the informal housing supply acts in a countervailing manner to the previously documented effect of geographic constraints in the housing supply.

We introduce two novel instruments for the housing supply curve, one based on the federal land holdings and another on the rate of natural increase of the adult population in the city. The latter can be used more broadly even in settings in which informality is not a concern as in most developed world. The federal land holdings instrument is a good predictor of informality in Brazil and could be further explored in other studies about housing informality. However it remains an open question if it could be applied more generally to other developing countries.

There is still much to be learned about informal housing market and its relation to the city's overall development. We hope this study of housing informality and the housing supply may open new paths for understanding the effects of different urban interventions (*e.g.*, infrastructure, public essential services, transportation, and real estate market regulation) in the developing world.

For instance, our model highlights a location-quality trade-off, taking as given the transportation network. As commonly described in developing countries, households in informal settlements can benefit from locations closer to economic activity and job opportunities, but face associated costs of potential eviction and inadequate urban infrastructure. It is plausible that improved transportation networks ease this trade-off and facilitate formal settlements away from city centers, in a cost-effective and welfare-improving manner. We leave this and other normative topics for future work.

References

Alonso, W. (1964). *Location and land use.* Cambridge, Massachusetts: Harvard University Press.

Alves, G. (2021). Slum growth in brazilian cities. *Journal of Urban Economics 122*, 103327.

Andrews, I., J. H. Stock, and L. Sun (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics 11*, 727–753.

Baum-Snow, N. and L. Han (2021). The microgeography of housing supply. Technical report, Working Paper.

Brueckner, J. K. (1987). The Structure of Urban Equilibria: A Unified Treatment of the Muth-Mills Model. *Handbook of Regional and Urban Economics 2*, 821–845.

Brueckner, J. K. (2013). Urban squatting with rent-seeking organizers. *Regional Science and Urban Economics 43*, 561–569.

Brueckner, J. K. and H. Selod (2009). A Theory of Urban Squatting and Land-Tenure Formalization in Developing Countries. *American Economic Journal: Economic Policy 1*(1), 28–51.

Cai, Y., H. Selod, and J. Steinbuks (2018, May). Urbanization and land property rights. *Regional Science and Urban Economics 70*, 246–257.

Cavalcanti, T., D. Da Mata, and M. Santos (2019). On the determinants of slum formation. *The Economic Journal 129*(621), 1971–1991.

Celhay, P. and R. Undurraga (2019). Location preferences and slums formation: Evidence from a panel of residence histories. Technical report, Colégio de México Working Paper.

Chaney, T., D. Sraer, and D. Thesmar (2012). The Collateral Channel: How Real Estate Shocks Affect Corporate Investment. *American Economic Review 102*(6), 2381–2409.

Chauvin, J. P., E. Glaeser, Y. Ma, and K. Tobio (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. Urbanization in Developing Countries: Past and Present.

Combes, P.-P., G. Duranton, and L. Gobillon (2019). The Costs of Agglomeration: House and Land Prices in French Cities. *The Review of Economic Studies 86*(4), 1556–1589.

Da Mata, D., S. V. Lall, and H. G. Wang (2008). Do Urban Land Regulations Influence Slum Formation? Evidence from Brazilian Cities. Technical report, IPEA.

Diamond, R. (2016). The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000. *American Economic Review 106*(3), 479–524.

Diamond, R. (2017). Housing Supply Elasticity and Rent Extraction by State and Local Governments. *American Economic Journal: Economic Policy 9*(1), 74–111.

Ferreira, P. C., A. Monge-Naranjo, and L. T. d. M. Pereira (2016). Of Cities and Slums. Technical Report 22, Federal Reserve Bank of St. Louis.

Flood, J. (2006). Secure Tenure Survey Final Report. Technical report, Urban Growth Management Initiative.

Guedes, R. B. (2020). *Essays on housing markets: evidences from Brazil*. Ph. D. thesis, FGV EPGE.

Gyourko, J., C. Mayer, and T. Sinai (2013). Superstar Cities. *American Economic Journal: Economic Policy 5*(4), 167–199.

Hansen, B. (2022). *Econometrics*. Princeton University Press.

Harari, M. (2020). Cities in bad shape: Urban geometry in india. *American Economic Review 110*(8), 2377–2421.

Henderson, J. V., T. Regan, and A. J. Venables (2021, May). Building the City: From Slums to a Modern Metropolis. *Review of Economic Studies 88*(3), 1157–1192.

Hilber, C. A. L. and W. Vermeulen (2015, 06). The Impact of Supply Constraints on House Prices in England. *Economic Journal 126*(591), 358–405.

IBGE (2011). *Censo Demografico 2010 - Aglomerado Subnormais*. IBGE.

Jimenez, E. (1985). Urban squatting and community organization in developing countries. *Journal of Public Economics 27*(1), 69–92.

Malpezzi, S. and D. Maclennan (2001, September). The Long-Run Price Elasticity of Supply of New Residential Construction in the United States and the United Kingdom. *Journal of Housing Economics 10*, 278–306.

Marx, B., T. Stoker, and T. Suri (2013). The Economics of Slums in the Developing World. *Journal of Economic Perspectives 27*(4), 187–210.

Mation, L. F., V. G. Nadalin, and C. Krause (2014). Favelização no brasil entre 2000 e 2010: Resultados de uma classificação comparável. Technical report, Texto para Discussão, No. 2009, Instituto de Pesquisa Econômica Aplicada.

Mian, A. and A. Sufi (2011). House Prices, Home Equity–Based Borrowing, and the US Household Leverage Crisis. *American Economic Review 101*(5), 2132–2156.

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

Muth, R. (1969). *Cities and housing.* Chicago, University of Chicago Press.

Nadalin, V. G. and L. Mation (2018). Localização interurbana das Favelas Brasileiras: o papel dos fatores Geográficos. Technical report, IPEA.

Oikarinena, E., R. Peltolab, and E. Valtonen (2015). Regional variation in the elasticity of supply of housing, and its determinants: The case of a small sparsely populated country. *Regional Science and Urban Economics 50*, 18–30.

Robb, A. M. and D. T. Robinson (2014). The Capital Structure Decisions of New Firms. *The Review of Financial Studies 27*(1), 153–179.

Saiz, A. (2010). The Geographic Determinants of Housing Supply. *Quarterly Journal of Economics 125*(3), 1253–1296.

Secretaria do Patrimônio da União (2010). *Balanço de Gestão 2003-2010.*

Shah, N. (2014). Squatting on Government Land. *Journal of Regional Science 54*(1), 144–136.

Smolka, M. O. and C. Biderman (2011). Housing informality: An economist's perspective on urban planning. In *The Oxford handbook of urban economics and planning*. Oxford University Press.

Stock, J. H., M. Yogo, et al. (2005). Testing for weak instruments in linear IV regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg 80*(4.2), 1.

Wang, S., S. H. Chan, and B. Xu (2012). The Estimation and Determinants of the Price Elasticity of Housing Supply: Evidence from China. *Journal of Real Estate Research 34*(3), 311–344.

A Proof of Theorem 2

Proof. First, note that

$$IHR_k = G\left(\frac{\Phi_{i,k}}{\Phi_{f,k}}\right),$$

where $G : \mathbb{R} \to \mathbb{R}$, in which $G(x) = \frac{1-\Omega_k}{\Omega_k}x^2 / \left(1 + \frac{1-\Omega_k}{\Omega_k}x^2\right)$. Notice that

$$G'(x) = \frac{2\frac{1-\Omega_k}{\Omega_k}x}{\left(1 + \frac{1-\Omega_k}{\Omega_k}x^2\right)^2} > 0. \tag{A.1}$$

Next, notice that

$$\frac{\Phi_{i,k}}{\Phi_{f,k}} = \frac{\Phi_{i,k}}{\Phi_{i,k} + \phi_k},$$

which by (14) is increasing in H_k and decreasing in Λ_k. This establishes (i) and (ii). It also follows from (14) that

$$\frac{d\Phi_{i,k}}{d\Omega_k} = -\frac{1}{2}\frac{(1-2\Omega_k)}{\Phi_{i,k} + \Omega_k \phi_k} - \phi_k.$$

So, $\Omega_k < \frac{1}{2}$ is sufficient to ensure that both $\Phi_{i,k}$ and IHR_k are increasing in $(1-\Omega_k)$. In order to establish (iv), first note that

$$\frac{\Phi_{i,k}}{\Phi_{f,k}} = \frac{\sqrt{\frac{H_k}{\pi\Lambda_k} - \Omega_k(1-\Omega_k)\phi_k^2} - \Omega_k\phi_k}{\sqrt{\frac{H_k}{\pi\Lambda_k} - \Omega_k(1-\Omega_k)\phi_k^2} + (1-\Omega_k)\phi_k}.$$

Now let

$$A := \sqrt{\frac{H_k}{\pi\Lambda_k} - \Omega_k(1-\Omega_k)\phi_k^2}$$

and

$$A' := -\frac{\Omega_k(1-\Omega_k)\phi_k}{A} < 0$$

Then, after some algebra, the derivative of the ratio of radii with respect to ϕ_k is

$$\frac{d\left(\frac{\Phi_{i,k}}{\Phi_{f,k}}\right)}{d\phi_k} = \frac{A'\phi_k - A}{(A + (1-\Omega_k)\phi_k)^2} < 0.$$

Finally for part (iii), notice that

$$\frac{dIHR_k}{d\phi_k} = \underbrace{\frac{dIHR_k}{d\left(\frac{\Phi_{i,k}}{\Phi_{f,k}}\right)}}_{+} \cdot \underbrace{\frac{d\left(\frac{\Phi_{i,k}}{\Phi_{f,k}}\right)}{d\phi_k}}_{-} \cdot \underbrace{\frac{d\phi_k}{d\lambda_k}}_{+} < 0,$$

where the sign of the second term was just established, the sign of the first term follows from (A.1), and the sign of the third term follows directly from (10). ■

B From the model to the empirical specification

The elasticity of the average rent with respect to population changes (i.e., a demand shift) can be written, after differentiating equation 17, as

$$\beta_k = \frac{d\overline{r}_f}{dh_k}\frac{1}{\overline{r}} + IHR_k \frac{d\left[\overline{r}_{i,k} - \overline{r}_{f,k}\right]}{dh_k}\frac{1}{\overline{r}} + \frac{dIHR_k}{dh_k}\frac{\left[\overline{r}_{i,k} - \overline{r}_{f,k}\right]}{\overline{r}}, \tag{B.1}$$

where $h_k := \ln H_k$.

Following a common property of circular cities, mean rents within each of the formal and informal housing sectors is given by the rent computed at 2/3 of the longest distance to the CBD within that sector. Then, from equation 2:

$$\overline{r}_{j,k} = r_{f,k}(0) - \frac{2}{3}t\Phi_{j,k} - \mathbb{1}_j \Psi_k, \text{ for } j \in \{i, f\}. \tag{B.2}$$

Therefore, given equations (8) through (10), the difference between average formal and informal rents does not respond to population changes, as

$$\overline{r}_{f,k} - \overline{r}_{i,k} = \frac{1}{3}\Psi_k - \frac{2}{3}(\lambda_k \cdot CC).$$

This ensures that the second term in equation B.1 vanishes. Now notice that from equation (12), we can express the informal household share as a function of h_k:

$$IHR_k(h_k) = \frac{\pi\Lambda_k(1-\Omega_k)\Phi_{i,k}(h_k)^2}{e^{h_k}}.$$

The informal household ratio elasticity with respect to city size is then

$$\frac{IHR_k^{'}(h_k)}{IHR_k(h_k)} = 2\epsilon_{i,k} - 1,$$

where $\epsilon_{i,k} \equiv \frac{\Phi_{i,k}^{'}(h_k)}{\Phi_{i,k}(h_k)}$ is the (informal sector) urban sprawl elasticity. Notice that this notion of sprawl elasticity is linked to the overall sprawl elasticity by equation 10. Therefore, the model delivers a novel connection between the elasticity of the informal household ratio and urban sprawl.

As a consequence, equation B.1 becomes

$$\beta_k = \frac{d \ln \overline{r}_{f,k}}{dh_k} \frac{\overline{r}_{f,k}}{\overline{r}_k} + (2\epsilon_{i,k} - 1) \frac{\left[\overline{r}_{i,k} - \overline{r}_{f,k}\right]}{\overline{r_k}} IHR_k \tag{B.3}$$

$$= \frac{d \ln \overline{r}_{f,k}}{dh_k} \frac{\overline{r}_{f,k}}{\overline{r}_k} + \Gamma_k^{SLUM} IHR_k, \tag{B.4}$$

for $\Gamma_k^{SLUM} := (2\epsilon_{i,k} - 1) \frac{\left[\overline{r}_{i,k} - \overline{r}_{f,k}\right]}{\overline{r_k}}$.

The supply elasticity in equation B.4 is decomposed in non-linear functions of primitive city characteristics, such as land unavailability, the share of dormant land, and eviction hazard rates.

First notice that a particular case of our model is the situation in which no vacant land enables informal occupation. That is, $\Omega_k = 1$ and we are back to the case studied by Saiz (2010), in which $\overline{r}_k = \overline{r}_{f,k}$ and only the first term in B.4 is present. In this case, only geographic constraints shape the supply elasticity.

Still, beyond that case, approximation arguments enable some additional conclusions. For example, for cities with small degrees of informality, i.e., $\Phi_{i,k} \approx 0$, we can show that $\epsilon_{i,k} > 1/2$ and $\frac{\overline{r}_{f,k}}{\overline{r}_k} \approx 1$. Then, as long as average informal rents are smaller than average formal rents, $\Gamma_k^{SLUM} < 0$ and cities with a higher informal household ratio exhibit a lower inverse supply elasticity.

Equation B.4 implies a cross-sectional relationship between informal housing prevalence and the inverse house supply elasticity. Our empirical focus is on this relationship. For instance, are cities in which eviction is harder or more land is left vacant, also cities in which a demand expansion leads to less pressure on prices?

While the relationship implied by Eq. B.4 is non-linear, we approximate it, up to first order, with expression 19. Notice that the coefficients β^{LAND} and β^{SLUM} describe, up to first order in the cross-section of cities, the effects land supply and equilibrium informality in the observed inverse supply elasticity. [24]

[24]Also, β^{SLUM} accounts for all the cross-sectional relationship between IHR_k and β_k. This includes effects through the response of formal rents, in the first term of equation B.4, and not only Γ_k^{SLUM}.

C Additional Tables

Table C.1: Hedonic Model

	(1)
Log(Rent-per-room)	OLS
House	–0.406***
	(0.002)
Electricity	0.620***
	(0.028)
Water pipe	0.181***
	(0.004)
Water net	0.058***
	(0.004)
Sewerage network	0.385***
	(0.002)
Septic tank	0.169***
	(0.003)
Masonry wall	0.331***
	(0.012)
Wood wall	0.102***
	(0.012)
Constant	3.096***
	(0.030)
N	909,052
R^2	0.175

Note: The dependent variable is log of rent-per-room. House is a dummy for households occupying an entire building. Water pipe and water net are houses with internal water pipes and connected to a water supply network respectively. Sewerage and septic tank are houses connected to a sewerage network and with septic tank respectively. Masonry wall and wood wall indicates the material used in the walls of the building. *, **, and *** indicates significance at the 10%,5%, and 1% levels, respectively. Standard errors in parentheses.

Table C.2: First-stage regressions

	ΔlnH_k	Unav. Land × ΔlnH_k	% Slums × ΔlnH_k	ΔlnH_k	Unav. Land × ΔlnH_k	% Slums × ΔlnH_k
	MSAs with Slums			All MSAs		
	(1)	(2)	(3)	(4)	(5)	(6)
ΔPredicted households ($\Delta lnPD_k$)	0.710***	-0.210**	0.053	0.494***	-0.132**	0.015
	(0.247)	(0.094)	(0.071)	(0.199)	(0.069)	(0.049)
Unav.30% 10km $\cdot \Delta lnPD_k$	-0.055	1.208***	0.199***	0.358	1.336***	0.213***
	(0.205)	(0.120)	(0.063)	(0.241)	(0.132)	(0.048)
%Public Land $\cdot \Delta lnPD_k$	0.072**	0.045*	0.030*	0.048*	0.029	0.035**
	(0.031)	(0.027)	(0.017)	(0.028)	(0.024)	(0.015)
Predicted households 2010 $\cdot \Delta lnPD_k$	-0.068**	-0.034**	0.020	-0.098**	-0.031	0.042*
	(0.031)	(0.016)	(0.015)	(0.048)	(0.020)	(0.021)
N	91	91	91	185	185	185
R^2	0.920	0.879	0.638	0.861	0.819	0.515
F-stat (excluded instruments)	6.74	37.17	4.82	10.65	51.26	9.41
SW underid test (p-values)	0.0004	0.0005	0.0012	0.0000	0.0000	0.0009

Note: Robust standard errors are in parentheses. This table reports 2SLS for each of the three first stage equations (22) for the sample of MSAs with slums (1-3) and for the full sample (4-6). Regional Dummies are South, Southeast, Center-West and North/Northeast. "% Public Land" is area of federal public properties without building divided by the MSA urban area. "Predicted households 2010" is the projection for 2010 population considering only mortality rates and population age distribution of each MSA in 1991. *, **, and *** indicates significance at the 10%,5%, and 1% levels, respectively. Regional dummies in all specifications. Heteroskedasticity-robust F test statistic for excluded instruments. Sanderson-Windmeijer p-values for the null that a particular endogenous regressor is unidentified.

Table C.3: Housing Quantity Instruments: Sample of all MSAs

Δ Households (ΔlnH_k)	(1)	(2)
Δ Predicted households ($\Delta lnPD_k$)	0.602***	0.663***
	(0.199)	(0.209)
Regional Dummies	No	Yes
R^2	0.079	0.856

Note: Robust standard errors are in parentheses. Sample of all MSAs (185 MSAs). The Demographic Instrument estimate the RNI of population over 20yo between 1991 and 2010 using mortality rates and age pyramid. Regional Dummies are South, Southeast, Center-West and North/Northeast. *, **, and *** indicates significance at the 10%,5%, and 1% levels, respectively.

Table C.4: Slums Instruments: Sample of all MSAs

% Slums/Households (IHR)	(1)	(2)
% Public Land†	0.017**	0.014*
	(0.006)	(0.008)
Predicted Households 2010	0.016**	0.016**
	(0.006)	(0.007)
Unav.30% 10 km	0.164***	0.156***
	(0.037)	(0.036)
Regional Dummies	No	Yes
R^2	0.391	0.557

Note: Robust standard errors are in parentheses. Sample of all MSAs (185 MSAs). Regional Dummies are South, Southeast, Center-West and North/Northeast. † Area of federal public properties without building divided by the MSA urban area. Predicted Households 2010 is the projection for 2010 population considering only regional mortality rates and population age distribution of each MSA in 1991. *, **, and *** indicates significance at the 10%,5%, and 1% levels, respectively.

Table C.5: **Table C.5:** Supply Curve 1991-2010: All MSAs

Δ*Rent index* $(\Delta ln(\tilde{r}_k))$	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)
ΔlnH_k	–0.297***	–0.306***	–0.292***	0.555	–0.220	0.354
	(0.091)	(0.085)	(0.086)	(0.610)	(0.281)	(0.464)
$Unav.30\%10km \cdot \Delta lnH_k$	0.056	0.003	0.005	–0.232	0.477*	0.270
	(0.119)	(0.162)	(0.161)	(0.249)	(0.247)	(0.385)
$IHR_k \cdot \Delta lnH_k$		0.181	0.281		–2.575***	–2.406*
		(0.451)	(0.435)		(0.983)	(1.235)
Regional Dummies	Yes	No	Yes	Yes	No	Yes
AR-p				0.5225	0.0000	0.0004
KP-p				0.0030	0.0264	0.0017
Overid.-p				Just-ID	0.6614	0.2080

Note: Heteroskedasticity robust standard errors in parentheses. This table displays estimates of the supply equation (20) for the sample of all MSAs (185 MSAs). The notation Δ before the variable means change between 1991 and 2010. The dependent variable is change in median rent-per-room index in all specifications. ΔlnH_k is change in the log of the number of households. IHR_k is the percentage of informal houses in 2010. $Unav.30\%10km$ is the share of unavailable land ($1-\Lambda_k$, in eq. (20)), in which we use a 10-km radius from the CBD and consider all land above 30% inclination as unavailable. AR-p indicates p-values for the Anderson-Rubin test of the null that all coefficients on endogenous variables are zero. KP-p indicates p-values for the Kleibergen-Paap test of the null of underidentification. Overid-p indicates the p-value for the Sargan-Hansen test of the null that instruments are valid. *, **, and *** indicates significance at the 10%,5%, and 1% levels, respectively.

Table C.6: Supply Curve 1991-2010: Non-coastal MSAs with at least one slum

ΔRent index ($\Delta ln(\tilde{r}_k)$)	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)
ΔlnH_k	−0.028	−0.122	−0.044	−0.189	−0.183	−0.145
	(0.142)	(0.145)	(0.151)	(0.609)	(0.253)	(0.565)
$Unav.30\%10km \cdot \Delta lnH_k$	0.296**	0.333**	0.242*	0.251**	0.501***	0.507***
	(0.132)	(0.143)	(0.141)	(0.121)	(0.157)	(0.163)
$IHR_k \cdot \Delta lnH_k$		0.109	0.253		−1.058*	−1.321**
		(0.650)	(0.612)		(0.554)	(0.584)
Regional Dummies	Yes	No	Yes	Yes	No	Yes
AR-p				0.1255	0.0004	0.0082
KP-p				0.0069	0.2832	0.1136
Overid.-p				Just-ID	0.6388	0.8381

Note: Heteroskedasticity robust standard errors in parentheses. This table displays estimates of the supply equation (20) restricting to the sample of non-coastal MSAs (60 MSAs). The notation Δ before the variable means change between 1991 and 2010. The dependent variable is change in median price-per-room index in all specifications. ΔlnH_k is change in the log of the number of households. IHR_k is the percentage of informal houses in 2010. $Unav.30\%10km$ is the share of unavailable land ($1-\Lambda_k$, in eq. (20)), in which we use a 10-km radius from the CBD and consider all land above 30% inclination as unavailable. AR-p indicates p-values for the Anderson-Rubin test of the null that all coefficients on endogenous variables are zero. KP-p indicates p-values for the Kleibergen-Paap test of the null of underidentification. Overid-p indicates the p-value for the Sargan-Hansen test of the null that instruments are valid. *, **, and *** indicates significance at the 10%,5%, and 1% levels, respectively.

Table C.7: Supply Curve 1991-2010: MSAs with at least one slum, 15% 50km geographical constraint

ΔRent index ($\Delta ln(\tilde{r}_k)$)	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)
ΔlnH_k	–0.060	–0.189	–0.055	0.424	0.337	0.622
	(0.178)	(0.151)	(0.177)	(0.623)	(0.465)	(0.591)
$Unav.15\%50km \cdot \Delta lnH_k$	0.169	0.387***	0.181	0.014	0.590***	0.486
	(0.166)	(0.134)	(0.182)	(0.200)	(0.215)	(0.310)
$IHR_k \cdot \Delta lnH_k$		–0.409	–0.087		–3.395**	–3.674*
		(0.274)	(0.290)		(1.421)	(1.953)
Regional Dummies	Yes	No	Yes	Yes	No	Yes
AR-p				0.5265	0.0000	0.0018
KP-p				0.0038	0.0888	0.0256
Overid.-p				Just-ID	0.4988	0.4113

Note: Heteroskedasticity robust standard errors in parentheses. This table displays estimates of the supply equation (20) using Saiz (2010) definition of geographical constraints for MSAs with slums (91 MSAs). The notation Δ before the variable means change between 1991 and 2010. The dependent variable is change in median price-per-room index in all specifications. ΔlnH_k is change in the log of the number of households. IHR_k is the percentage of informal houses in 2010. $Unav.15\%50km$ is the share of unavailable land ($1-\Lambda_k$, in eq. (20)), in which we use a 50-km radius from the CBD and consider all land above 15% inclination as unavailable. AR-p indicates p-values for the Anderson-Rubin test of the null that all coefficients on endogenous variables are zero. KP-p indicates p-values for the Kleibergen-Paap test of the null of underidentification. Overid-p indicates the p-value for the Sargan-Hansen test of the null that instruments are valid. *, **, and *** indicates significance at the 10%,5%, and 1% levels, respectively.

Table C.8: Supply Curve 1991-2010: Limited Information Maximum Likelihood (LIML) estimates

ΔRent index ($\Delta ln(\tilde{r}_k)$)	LIML (1)	LIML (2)	LIML (3)
ΔlnH_k	0.393	0.025	0.615
	(0.589)	(0.362)	(0.631)
$Unav.30\%10km \cdot \Delta lnH_k$	0.150	0.843***	1.005**
	(0.178)	(0.275)	(0.420)
$IHR_k \cdot \Delta lnH_k$		−3.625***	−4.795**
		(1.280)	(2.076)
Regional Dummies	Yes	No	Yes
AR-p	0.3506	0.0000	0.0004
KP-p	0.0035	0.0869	0.0315
Overid.-p	Just-ID	0.5267	0.2714

Note: Heteroskedasticity robust standard errors in parentheses. This table displays estimates of the supply equation (20) by LIML using the sample of MSAs with slums (91 MSAs). The notation Δ before the variable means change between 1991 and 2010. The dependent variable is change in median rent-per-room index in all specifications. ΔlnH_k is change in the log of the number of households. IHR_k is the percentage of informal houses in 2010. $Unav.30\%10km$ is the share of unavailable land ($1-\Lambda_k$, in eq. (20)), in which we use a 10-km radius from the CBD and consider all land above 30% inclination as unavailable. AR-p indicates p-values for the Anderson-Rubin test of the null that all coefficients on endogenous variables are zero. KP-p indicates p-values for the Kleibergen-Paap test of the null of underidentification. Overid-p indicates the p-value for the Sargan-Hansen test of the null that instruments are valid. *, **, and *** indicates significance at the 10%,5%, and 1% levels, respectively.

D Simulated price trajectories

As an illustrative experiment, we use the cross-section of inverse elasticity estimates to simulate price increases that would be expected given future natural population growth in Brazilian cities. We simulate the increase in population growth in each city between 2010 and 2030 based on the RNI.

We compute price increases that would accommodate the increase in demand growth under the assumption of perfectly inelastic demand. That is, we do not allow for the demand side to respond, for instance, through changes in the birth rate, migration, and adult cohabitation.[25] This is indeed a strong assumption, but it serves our purposes of illustrating the dispersion in price increases produced by the cross-city variation in supply-side elasticities. These results can nevertheless be interpreted as an upper bound on the equilibrium price changes absent of other demand shocks.

We remove 7 MSAs from this experiment due to negative point estimates for the inverse supply elasticities. São Luís (MA), Teresina (PI), Recife (PE), São Mateus (ES), Angra dos Reis (RJ) and Resende (RJ) have estimated inverse supply elasticities between -0.14 and -0.08, which are values close to zero considering standard errors around 0.45. In addition, Belém (PA) has an inverse elasticity of -1.22 as the result of 54% informal household ratio, an outlier in the dataset.

Given an inelastic demand, the predicted price increase in a given city k is just:

$$\Delta ln\widetilde{P_k}^{2010-2030} = \hat{\xi}_k^S \cdot \Delta lnPD_k^{2010-2030},$$

where $\Delta lnPD_k^{2010-2030}$ is the expected future increase in population. Table D.1 presents the growth in the number of households and the corresponding price change.

[25] Guedes (2020, chap. 3) uses a similar dataset to discuss the effects of rents on birth rates, marriage, and adult cohabitation in Brazil.

Table D.1: Inverse Supply Elasticity, RNI, and Price Growth between 2010 and 2030†

MSAs	Inv. Elast.	SE	ΔSupply	ΔPrice	MSAs	Inv. Elast.	SE	ΔSupply	ΔPrice
Nova Friburgo*/RJ	0.99	0.53	16%	16%	Ribeirão Preto/SP	0.39	0.48	22%	9%
Florianópolis/SC	0.86	0.50	25%	22%	João Pessoa/PB	0.38	0.45	30%	12%
Rio Grande*/RS	0.81	0.50	23%	19%	Jundiaí/SP	0.38	0.47	25%	9%
Itajaí - B. Camboriú/SC	0.80	0.50	31%	25%	Fortaleza/CE	0.38	0.44	33%	13%
Petrópolis/RJ	0.79	0.48	19%	15%	Mossoró*/RN	0.37	0.48	31%	12%
Juiz de Fora/MG	0.74	0.50	20%	15%	Vitória/ES	0.37	0.45	30%	11%
Colatina*/ES	0.66	0.48	24%	16%	C. Itapemirim*/ES	0.36	0.44	27%	10%
Pelotas/RS	0.65	0.49	20%	13%	Montes Claros*/MG	0.36	0.47	35%	13%
Parintins*/AM	0.63	0.46	72%	46%	Guarapari*/ES	0.35	0.44	32%	11%
Itabira*/MG	0.62	0.48	28%	18%	Londrina/PR	0.35	0.48	24%	8%
Cametá*/PA	0.61	0.47	60%	37%	Campos dos G./RJ	0.34	0.48	28%	9%
Gov. Valadares*/MG	0.61	0.48	31%	19%	Piracicaba/SP	0.33	0.48	22%	7%
Tramandaí - Osório/RS	0.58	0.45	25%	14%	Caruaru*/PE	0.33	0.47	32%	11%
Corumbá/Brasil	0.56	0.47	47%	27%	Novo Hamburgo/RS	0.32	0.47	28%	9%
Atibaia/SP	0.56	0.49	24%	13%	Ponta Grossa/PR	0.32	0.48	36%	11%
Maceió/AL	0.55	0.45	37%	21%	Ipatinga/MG	0.32	0.45	31%	10%
Porto Alegre/RS	0.53	0.46	23%	12%	Cabo Frio/RJ	0.31	0.43	30%	9%
Joinville/SC	0.53	0.48	31%	17%	Brasília/DF	0.31	0.47	39%	12%
Marília/SP	0.51	0.48	19%	9%	Aracaju/SE	0.28	0.45	36%	10%
Taubaté/SP	0.49	0.49	27%	13%	Araguaína*/TO	0.28	0.47	46%	13%
Itabuna*/BA	0.49	0.49	29%	14%	Macapá/AP	0.26	0.43	60%	15%
Natal/RN	0.48	0.45	30%	15%	Parauapebas*/PA	0.25	0.46	60%	15%
Boa Vista*/RR	0.48	0.49	54%	26%	Baixada Santista/SP	0.23	0.43	23%	5%
Blumenau/SC	0.47	0.46	27%	13%	Ilhéus*/BA	0.20	0.42	34%	7%
Bento Gonçalves/RS	0.47	0.46	18%	8%	Curitiba/PR	0.20	0.47	30%	6%
Americana/SP	0.47	0.49	20%	9%	Campina Grande/PB	0.20	0.47	31%	6%
S. J. dos Campos/SP	0.46	0.48	28%	13%	Cuiabá/MT	0.20	0.47	36%	7%
Itu - Salto/SP	0.46	0.49	26%	12%	Teresópolis*/RJ	0.17	0.42	23%	4%
Campo Grande*/MS	0.45	0.49	31%	14%	Campinas/SP	0.14	0.47	24%	3%
Arapiraca*/AL	0.45	0.49	41%	18%	Linhares*/ES	0.11	0.46	38%	4%
Goiânia/GO	0.45	0.49	32%	15%	Belo Horizonte/MG	0.11	0.46	27%	3%
Passos/MG	0.45	0.48	25%	11%	Marabá*/PA	0.11	0.46	58%	6%
Umuarama/PR	0.45	0.49	23%	10%	Porto Velho/RO	0.11	0.46	46%	5%
Anápolis*/GO	0.44	0.49	32%	14%	Rio Branco*/AC	0.08	0.47	52%	4%
Foz do Iguaçu/PR	0.44	0.48	41%	18%	Manaus*/AM	0.08	0.45	50%	4%
Caxias do Sul/RS	0.44	0.46	24%	11%	São Paulo/SP	0.06	0.47	26%	2%
Rio de Janeiro/RJ	0.42	0.44	21%	9%	Salvador/BA	0.04	0.43	29%	1%
Santarém*/PA	0.42	0.43	54%	23%	Araruama/RJ	0.02	0.45	24%	0%
Paranaguá*/PR	0.42	0.45	41%	17%	Recife/PE	-0.08	0.45	-	-
Passo Fundo*/RS	0.42	0.48	26%	11%	Resende/RJ	-0.09	0.46	-	-
Bauru/SP	0.41	0.48	20%	8%	São Mateus*/ES	-0.09	0.47	-	-
Volta Redonda/RJ	0.40	0.45	22%	9%	Teresina/PI	-0.14	0.48	-	-
Juazeiro do Norte/CE	0.40	0.48	37%	15%	São Luís/MA	-0.14	0.45	-	-
Macaé/RJ	0.39	0.44	32%	12%	Angra dos Reis*/RJ	-0.14	0.44	-	-
Tubarão - Laguna/SC	0.39	0.47	22%	9%	Belém/PA	-1.22	0.71	-	-
Sorocaba/SP	0.39	0.48	26%	10%					

Note: †The rate of natural increase considers only the aging of people under 20 years old and the mortality rate of older cohorts to predict the population growth for citizens over 20 years old. *Isolated municipalities.