Can Big Data increase our knowledge of local rental markets? Estimating the cost of density with rents *

Guillaume Chapelle †and Jean-Benoît Eyméoud‡

Abstract

In this paper, we argue that the cost of agglomeration should preferably be measured with rents since the cost of housing based on prices is forward looking and might depend on parameters likely to vary with city size. As access to rental data is usually limited, we create a new data set regularly scraping two major French real estate websites. Comparing our data set with the French Housing survey only available at the department and regional level, we show that internet-based estimates are not biased as they do not systematically differ from surveys. We then use our data set to create a comparable rent measure for every urban area in France. We show that rent/price ratios are lower in large agglomerations resulting in a lower elasticity of housing cost with respect to city size when measured with rents instead of prices as in the seminal contribution of Combes, Duranton, and Gobillon (2018). This result is of particular importance when computing the net benefits of density which appear larger and also positive for renters.

JEL codes Housing Demand (R21), Housing Supply (R31), Large Data Sets: Modeling and Analysis (C55), Methodology for Collecting, Estimating, and Organizing Microeconomic Data - Data Access (C81)

*The authors acknowledge the support from ANR-11-LABX-0091 (LIEPP), ANR-11-IDEX-0005-02 and ANR-17-CE41-0008 (ECHOPPE). They also thank participants in the Large Open/Online Raw Dataset (LOORD) seminar in Sciences Po, the ECHOPPE seminar in Toulouse School of Economics and the European Meeting of the Urban Economics Association for their comments and questions. They are particularly grateful to Jean-Charles Bricongne, Julia Cagé, Pierre-Philippe Combes, Pierre Deschamps, Gilles Duranton, Laurent Gobillon, Morgane Laouennan, Philippe Martin, Joan Monras, Florian Oswald, Bruno Palier, Quentin Ramond, Claudia Senik, Michel Serafinelli, Claude Taffin, Corentin Trévien, Grégory Verdugo, Paul Vertier, Benjamin Vignolles and Etienne Wasmer for their helpful comments and discussions. All errors remain ours.

†Sciences Po, Department of Economics, LIEPP, Paris

‡Sciences Po, Department of Economics, LIEPP, Paris

1 Introduction

Cities are often presented as the equilibrium outcome between agglomeration economies and agglomeration costs (Fujita and Thisse (2002)). While positive agglomeration externalities have long been studied (Combes and Gobillon (2015)), few works document the costs of agglomeration. In a recent paper Combes, Duranton, and Gobillon (2018) carefully document the housing cost of agglomeration which is the product of the share of housing consumption and the elasticity of housing costs with respect to city size. While the share of housing consumption is measured with rents and imputed rents, the authors measure the elasticity of housing cost with respect to city size using land and housing prices. This methodological choice might not be innocuous if the parameters of the user costs such as local tax rates, price expectations or interest rates also vary with city size. Moreover, as prices are forward looking and influenced by expected capital gains (Glaeser and Nathanson (2017)), they also might diverge from their fundamental value. For example, Verbrugge (2008) and Garner and Verbrugge (2009) show that standard user cost approach computed from prices are not in line with the market rents while Ambrose, Eichholtz, and Lindenthal (2013) show that prices can diverge from their fundamentals during decades. It might thus be preferable to measure the elasticity of the housing cost with respect to city size with rents instead of prices. Nevertheless, if transactions prices are systematically recorded by the fiscal administration or solicitors, few data sets gather precise information on rents. In the meantime, housing surveys show that an increasing share of tenants find their accommodation thanks to real estate websites. Gathering millions of rental posts and refreshing them continuously, these websites offer a precise real time picture of the rental market. In this paper, we propose a new database built thanks to online ads and use it to estimate the cost of agglomeration based on rents instead of prices.

Between December 2015 and January 2018, we have periodically collected, cleaned and analyzed housing rental posts coming from the two largest French real estate websites. Each post provides the location of the housing good as well as its hedonic characteristics, offering the possibility to describe local housing markets and create local rent indices. First, we assess whether posted rents diverge from signed contracts observed in the French housing survey where data are available only at the department and regional level. We thus build two rent indices for French Regions and Departments based on signed contracts and online ads. Comparing our results, we show that no systematic bias appears between the two sources of data.

We then exploit our original data set to reproduce Combes, Duranton, and Gobillon (2018) and measure the relationship between urban density and housing cost using rents instead of prices. This exercise appears particularly important since the cost of housing and housing prices are not equivalent. Indeed, homeowners pay property taxes and interest rates but can benefits from capital grains. In our view, as the share of housing in expenditures is measured with market and imputed rents, the elasticity of housing cost with respect to city size might be measured with rents. Moreover, using rents instead of prices might mitigate problems related with the fact that prices are forward looking and capitalize expected phenomenon as city growth or other unobserved amenities. Our intermediate results seem to confirm the standard user cost approach as Rent/Price ratios appear negatively correlated with past price growth and positively correlated with the property tax rate. When measuring the cost of agglomeration with rents, we estimate an elasticity which is almost half the one measured with housing prices. This result appears particularly relevant when measuring the net benefits of density as in Ahlfeldt and Pietrostefani (2019). Using our estimates, we find that density also appears beneficial for renters. Our findings support the claim that cities are close to the equilibrium where urban costs are extremely close to agglomeration economies.

This paper belongs to the growing stream of literature which exploits user-generated content to study urban phenomenon. The closest contribution to our work is Boeing and Waddell (2016) who use Craiglist rental housing listings to study the rental housing market in the US. Other works using similar data such as Loberto, Luciani, and Pangallo (2018) in Italy or Bricongne, Pontuch, and Turrini (2017) for Europe usually focus on housing prices. We distinguish our dataset from previous works by focusing on rents and exploiting information from several websites. Similar datasets were also gathered to answer some particular research questions. For example, Mense, Michelsen, and Cholodilin (2017) develop a similar database based on one single website in order to assess the impact of rent control in Germany. Other studies as Loberto, Luciani, and Pangallo (2018) use online data provided by the websites without scraping. For example, Brülhart et al. (2017) use this kind of data to measure the housing supply and demand elasticities and to assess the tax incidence in Switzerland. Ads were also exploited in Basten, Von Ehrlich, and Lassmann (2017) to investigate spatial sorting on the Swiss market while Kholodilin, Mense, and Michelsen (2017) and Hyland, Lyons, and Lyons (2013) use online data to investigate the impact of energy efficiency labels on housing prices. Other papers exploit user generated content in order to investigate segregation in consumption as Davis et al. (2017) in New York or to investigate discrimination on short term rental markets (Edelman and Luca (2014) and Laouénan and Rathelot (2017)).

We also belong to the literature documenting the costs and benefits of density extensively reviewed in Ahlfeldt and Pietrostefani (2019). More specifically, we document carefully the housing cost of density following closely the methodology developed in Combes, Duranton, and Gobillon (2018). We also contribute to the literature documenting the spatial disparities between rents and prices as Himmelberg, Mayer, and Sinai (2005) who compute the evolution of the user cost for several Metropolitan Statistical Areas in the US or Gyourko, Mayer, and Sinai (2013) who document higher price income ratios in the most dynamic cities and neighborhoods in the US. We are also related to the literature on measuring quality of life in cities as exposed in Albouy (2008). Moreover, our findings on the relationship between urban prices and rents can also contribute to the growing literature documenting the value of urban land such as Davis and Heathcote (2007) and Albouy, Ehrlich, and Shin (2018).

Our contributions to the literature are the following. First, we argue that online data provide an accurate view of local rental market when compared with survey data. Second, our main application builds on previous empirical works on urban and housing economics and provides an accurate estimate of the elasticity between city size and housing cost based. This complements the recent work by Combes, Duranton, and Gobillon (2018) which focuses on prices documenting more precisely the difference between the cost of owning and housing prices. Our new estimates of the cost of density is of particular importance in order to compute its net effect as in Ahlfeldt and Pietrostefani (2019). By providing a new upper bound of the cost of agglomeration which can be up to 40%

lower from the previous estimates, we reinforce the claim that the net effect of density is positive. Our results also strongly support the idea that cities are in equilibrium as urban costs appear extremely close to agglomeration economies.

The paper is structured as follows: Section 2 discusses why the cost of agglomeration should be preferably measured with rents rather than with prices. Section 3 presents briefly the estimation framework developed in Combes, Duranton, and Gobillon (2018) to measure the cost of density. Section 4 presents our data set built to revisit the cost of density with a particular focus on the webscraping of online ads and its reliability to measure market rent. In Section 5, we estimate the cost of density based on rents instead of prices and discuss the implications of our new elasticity in the assessment of the benefits of density. Section 6 concludes.

2 Defining the cost of housing

In Poterba (1984), Verbrugge (2008), Himmelberg, Mayer, and Sinai (2005), and Hill and Syed (2016) housing cost for homeowners is is the user cost of housing capital. At the equilibrium, Himmelberg, Mayer, and Sinai (2005) writes the relationship between the cost of owning and renting as follows :

$$R_i = C_i \tag{1}$$

where R_i is the rent level in city i and C_i is the cost of homeownership. Adapting Himmelberg, Mayer, and Sinai (2005) to the French context and neglecting the transaction tax, the cost of home ownership (C_i) could be expressed as follows:

$$C_i = (\tau_i + r_i + \delta_i - g_i) \times P_i \tag{2}$$

where τ_i is the local property tax rate. r_i is the interest rate, δ_i is the depreciation rate and g_i is the growth rate of housing prices. In France homeowners cannot itemize interest rates (excepted between 2007 and 2011). In such a framework, the indifference between annual cost of owning and renting would imply that:

$$P_i = \frac{R_i}{\tau_i + r_i + \delta_i - g_i} \tag{3}$$

In this paper, we want to argue that the cost of agglomeration which is mainly the cost of housing should be preferably measured with rents. In their seminal work, Combes, Duranton, and Gobillon (2018) define the cost of agglomeration ϵ_N^{UC} as follows:

$$\epsilon_N^{UC} = s_E^h \times \epsilon_N^{\bar{P}} \tag{4}$$

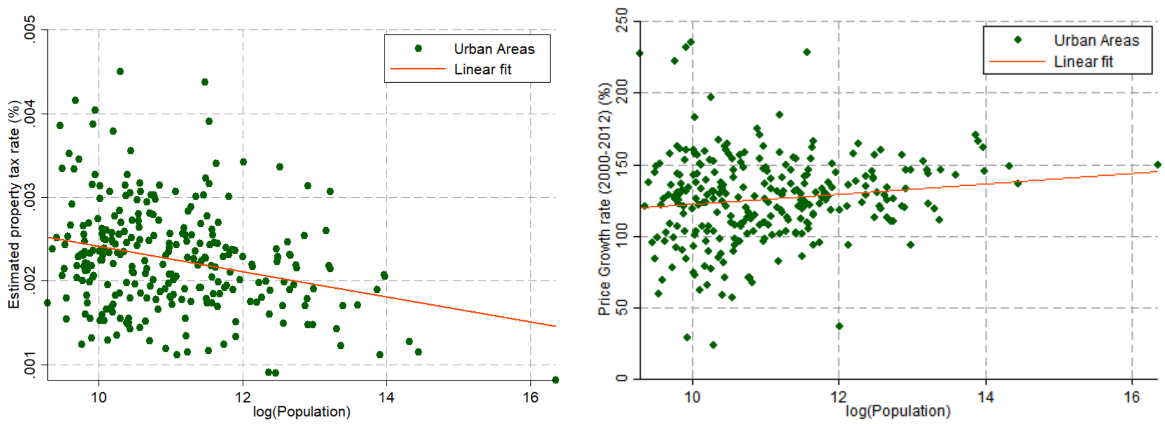
where s_E^h is the housing share in expenditure and $\epsilon^{\bar{P}_N}$ is the elasticity of housing cost with respect to city population and can be defined as follows :

$$\epsilon_N^{\bar{P}} = \frac{\partial C}{\partial Pop} = \frac{\partial R}{\partial Pop} \tag{5}$$

where Pop is the Population of the city. In their paper, the authors estimate $\epsilon^{\bar{P}_N}$ with land and housing prices while s_E^h is estimated with imputed rents and market rents[1].

To our understanding, equation 3 supports the view that $\epsilon_N^{\bar{P}}$ should be preferably measured with rents since many parameters might affect C_i disturbing the relationship between price and housing cost, i.e. $\frac{\partial C}{\partial Pop} \neq \frac{\partial P}{\partial Pop}$. Indeed, an estimate of the cost of agglomeration based on prices might be biased if denser place have different tax rates, interest rates, depreciation rates and expected growth rates. As housing prices indices account for the period of construction, it is unlikely that the depreciation rate would affect the results. On the other hand property tax and expected price growth might be a concern. If the property tax rate is negatively correlated with population and expected price growth is positively correlated with population, the authors might overestimate the cost of agglomeration. We illustrate our concerns in Figure 1. In Panel A), we estimate a property tax rate[2]. We measure for each urban area the average property tax per square meter and obtain a tax rate dividing the average tax by the price per square meter. This estimated tax rate appears negatively correlated with city size. Moreover, panel B) tries to approximate expected price growth with the growth rate between 2000 and 2012. Again, larger cities tend to have higher price growth which might influence the expected growth rates as stressed in Glaeser and Nathanson (2017). Two approaches might be used to deal with this concern. The first one would be to compute a user cost for each urban area following Himmelberg, Mayer, and Sinai (2005). However, assessing the expected growth rate remains a difficult task and often relies on strong assumptions. Moreover, access to local interest rates remains difficult. The second solution is to estimate the cost of density with rents which should remove these three concerns and provide us an unbiased upper bound for the cost of density provided that owners will buy a house as long as $C_i < R_i$. This should also improve the coherence between $\epsilon_N^{\bar{P}}$ and s_E^h.

Figure 1: Correlation between taxes, price growth and city size



(a) Estimated Property tax rate (b) Price growth

There are other reasons to prefer the use of rents instead of prices. First, several

1. s_E^h is estimated to be 0.159 in Combes, Duranton, and Gobillon (2018) and increases with city size. Ahlfeldt and Pietrostefani (2019) retains a share of housing expenditure of 0.33.

2. In practice the property tax does not depend on housing prices but is the product of a municipal rate applied on cadastral values (valeurs locatives) computed in 1970.

papers also emphasized that prices might diverge from their fundamental values. For example, Ambrose, Eichholtz, and Lindenthal (2013) shows that prices can remain far from the fundamental values over long period of time. Bonnet et al. (2014) also documents diverging trends between rents and prices in France and other developed economies. Such aggregate patterns can be partly explained in Verbrugge (2008) who shows that rents and standard user cost computed from housing prices might not always be used to understand the trade-off between renting and owning. Gallin (2008) and Campbell et al. (2009) also provide evidence that the rent price ratio allows to predict the future evolution of housing prices in line with the idea that rents is closer to the fundamental value of housing. Finally, a growing literature emphasize the role of expectation in housing price formation that tend to increase the likelihood of bubbles (Glaeser and Nathanson (2017, 2015))

Moreover, standard urban theory predicts that city growth might affect the rent/price ratio, Combes, Duranton, and Gobillon (2018) control for the annualized urban growth rate in some of their specifications. However, it is worth noting that the rent price ratio might have diverging patterns within cities at odd with the standard model. For example, Chapelle, Wasmer, and Bono (2017) or Halket, Nesheim, and Oswald (2015) document strong discrepancies between the rent and the price gradients both in Paris and London. These papers show that the rent gradient is much lower than housing prices which is at odd with the standard prediction of the monocentric model when cities are growing as exposed in DiPasquale and Wheaton (1996). As emphasized in Glaeser and Gyourko (2007), more works are thus required to understand the relationship between prices and rents. Because housing is both an investment and a consumption good, its rent appears as a better measure of its annualized cost.

Finally, one should keep in mind that the center of French Urban Areas as Paris combine high levels of income and low home-ownership rates. In the Urban area of Paris, the home-ownership rate is 47% which is 10 percentage points below the national average. More generally, only 53% of the households living in one of the 354 Urban areas own their main residence while 70% of these living in a rural municipality are homeowners. Moreover, Combes, Duranton, and Gobillon (2018) focus on the value of housing in the City Business District where home-ownership tends to be extremely low.

To summarize, as the relationship between rents and prices might be affected by several parameters and because housing prices might diverge from their fundamental values, it appears reasonable to estimate the elasticity between housing costs and population using rents when available.

3 Estimating the cost of density

This section introduces our methodology to estimate the relationship between the city size and the cost of housing in the city center. We rely on the methodology developed in Combes, Duranton, and Gobillon (2018) based on housing prices.

3.1 A comparable hedonic index for each city center

First, we estimate an hedonic index of the rent for each French urban municipalities:

$$ln(r_{i,s}) = ln(r^{ref}_{m(i)}) + X_i\beta_s + u_i \quad (6)$$

Where $ln(r_{i,s})$ is the rent per square meter of unit i in strata s, $ln(r^{ref}_{m(i)})$ is the hedonic index of the municipality m where the unit is located, X_i is a vector of hedonic characteristics of the unit (surface, number of rooms, presence of other amenities (furnished, garden, balcony, garage, elevator)) and β_s is the vector of corrective coefficients which are allowed to vary between strata. We estimate our hedonic model separately for each department. One could fear that online ads might be biased by negotiations. However, the next section tends to support that these type of hedonic indices do not present any particular bias when compared with signed contracts.

These hedonic indices are then used to recover comparable rents in the City center ($C^r_{c(m)}$) which is a measure of the urban cost. Measuring the rent in the city center, which is the weighted average of the location of all Jobs, allows to neglect transport costs. Following Combes, Duranton, and Gobillon (2018), we estimate the following equation:

$$ln(r_m) = C^r_{c(m)} - \delta^r_{c(m)} \times ln(D_m) + X_m\alpha^r + \mu_m \quad (7)$$

where $ln(r_m)$ are the hedonic indices resulting from the estimation of the preceding equation while $C^r_{c(m)}$ is an urban area fixed effect. $\delta^r_{c(m)}$ measures the rent gradient of each urban area. The CBD is measured as the barycenter of employment of the urban area following Combes, Duranton, and Gobillon (2018). $X_m\alpha^r$ are municipalities controls as the amenities (high schools, doctors, hospitals, temperature, altitude) reproduced following the instructions in Combes, Duranton, and Gobillon (2018) with the exception of the standard deviation in income which is not publicly available and where sources are specified in Table 11. The estimated $C^r_{c(m)}$ are displayed in panel C) of Figure 6. It is worth noting that $C^r_{c(m)}$ can also directly inferred from our micro data including municipalities control instead of fixed effects and controlling for the distance from the CBD in the first equation. This does not change our results.

3.2 Identification of the impact of density

Finally, we estimate the elasticity of housing rent with respect to density estimating the following equation:

$$\hat{C}^r_c = \epsilon^{\bar{P}}_N \times ln(Population_c) + \gamma \times ln(Land\_Area_c) + Z_c\beta + \eta_c \quad (8)$$

where $\hat{C}^r_c$ are the urban area fixed effects which is the log rent in the city center. As we control for the surface of the urban area, the coefficient of $ln(Population)$ is our parameter of interest and depicts the elasticity of urban rent with respect to density ($\epsilon^{\bar{P}}_N$). Z_c is a set of control variable as the level of income and education in the urban area, the population growth and also some geological variables, the share of the area covered by buildings and their average height. As the estimate of $\epsilon^{\bar{P}}_N$ might be biased by unobserved variable and reverse causality, it is possible to exploit a set of instruments suggested in Combes, Duranton, and Gobillon (2018). The instruments are divided between two categories : Historical population and density (with data from the XIX century) and Natural

Amenities (Number of Hotel rooms, share of budget hotel rooms and the temperature in January). The two complementary sources of exogenous variation allows to perform over identification tests. Besides the data provided in the original study, we extend their series for all the urban areas in France using the historical data set on population gathered in Motte, Séguy, and Théré (2003). This data set is exhaustive contrary to the one used by the authors which covered only the population of large municipalities. We also improve their measure of temperature using the data provided in Hijmans et al. (2005) which allows to compute precisely the temperature in the Urban area and not at the department level. Our sources are summarized in Table 11

4 The Data

Nowadays a vast majority of private landlords or real estate agencies use internet to find tenants as illustrated in Table 1. Even if these channels do not constitute the whole market, as 22% of the tenants found their flat by alternate channels[3], one can think that we are able to observe the vast majority of the market. Ads posted online can thus be an interesting way to follow the rental sector dynamics.

Table 1: Method used to find a Flat in the rental sector (%)

	Not Furnished	Furnished	Total
Privately (ads on internet or Newspapers)	37	42	37
Real Estate Agency	41	22	39
by word of mouth	19	20	19
From the employer	1	3	2
Social Services	2	10	3
Others	0	3	1
Total	100	100	100

Source : Author's computation from the French Housing Survey 2013 (INSEE)
Households in the private rental sector installed for less than 4 years.

This source of information has special features. First, it allows to measure the level of the rent for new tenants who represent 18% of the Rental Sector as illustrated in Table 2. This fact is of particular importance given the regulation of the French Housing Market. Indeed, once the contract is signed, yearly revision of the rent level cannot exceed an official index : the Rent Revision Index[4] which is a price index from which tobacco and rental prices were removed. Such a regulation is defined as Rent Control of Type 2 in Arnott (1995). As a consequence, such transactions can only provide information about the change on the flow of rental housing as when observing housing price transaction. The rent we follow corresponds to the rent for new rentals while the rent index computed by National Agencies as the French Institut National de la Statistique et des

3. Namely 19% by word of mouth, 1% from the employer and 2% from social services.
4. Indice de Référence des Loyers (IRL)

Etudes Economiques (INSEE) corresponds to an index for the stock of rentals. Following new leases can provide us additional information.

Table 2: Time of occupation of the housing unit (%)

	Not Furnished	Furnished	Total
Less than 1 year	18	39	19
1 to 4 years	28	36	28
4 to 8 years	18	15	18
8 to 12 years	9	5	9
more than 12 years	27	5	26
Total	100	100	100

Source : Author's computation from EL 2013 (INSEE)
Households in the private rental sector

4.1 Scraping process

There are several rental websites in France. To get access to the biggest source of data we decided to focus on the two largest. The first has about half of its posts from landlords and half from real estate agents. The second has mostly his post from real estate agents. The information we want to extract consists of a set of posts that are available on the rental websites. Each post is a Volunteered Geographic Information (VGI) as it is both user generated and geolocated (Jiang and Thill (2015)). It has a unique identifier, pictures, a short text describing the offer and a standardized table presenting the most important characteristics as the surface, the number of rooms, the monthly rent or the type of contract (furnished or not). It is also localized thanks to the name of the municipality, a zip-code and a map indicating the geographic coordinates which can be more or less precise (city level, neighborhood or address). The non-structured part of the post (description) allows to identify key words in order to find additional information as the presence of an elevator, the floor, the amount of extra expenditures.

To get the data from the two websites, we use Python to create programs that mimic a web browser request. The first step consists in finding the Uniform Resource Locator (URL) of each post which concerns the rental market in France. In a second step, we extract the Hypertext Markup Language (HTML) of each page from the server. In a third step, we clean it and structure it so as to get a structured format for each post. Finally, we save the database in comma separated values format. Overall, the operation takes between 10 hours and two days depending on the Website and the period of time. We repeat the process of scraping every month for each website from December 2015 until January 2018 and end up with a database of 4.2 millions posts in the rental sector.

This dataset appears much more precise that existing surveys on rents. Indeed, the

French Housing Survey[5] or the Survey on Rents and Housing Expenditures[6] provide good quality data on the rental sector but have two drawbacks. First, they are only representative at the national level as they have a limited number of observations[7]. Thus, one cannot use them to follow the rental dynamics of a city or an urban area. On the other hand, they do not allow to follow the housing market on a monthly basis as they are published every four years (French Housing Survey) or every quarter (Survey on Rents and Housing expenditures). Moreover, each wave observe the stock of rental unit and not the flow. There are other rental data sets as the Observatoire des Loyers de l'Agglomération Parisienne (OLAP)[8], Connaître Les Loyers et Analyser les Marchés sur les Espaces Urbains et Ruraux (CLAMEUR)[9] and more recently the group of local observatories, Observatoires Locaux des Loyers (OLL)[10]. These data are usually focused on the largest urban areas and are also representative of the stock. Moreover, their datasets were not made public so far. Nevertheless as it corresponds to posted rents and not to signed contract, it appears important to assess whether a discrepancy exists between survey data and posted rents.

4.2 Cleaning the data and creation of the variables

The cleaning procedure starts by identifying the repeated posts which have the same identifier between each wave. We also identify similar posts between both sites using the post's description. We keep only one observation by post and keep the number of occurrences of the post. our approach is different from Loberto, Luciani, and Pangallo (2018) that use a Machine Learning algorithm to identify similar ads with different Identifier. In our approach, we generated a full set of variables describing each unit (see below) and consider that a unit with the same price and the same characteristics (rent, surface, number of rooms, amenities and geocoding) posted in the same month are duplicates. We keep only the posts that have a price per square meter which is strictly positive and lower than 100[11] and above 2. This procedure creates the final database that we describe in the next section. For each post, we compute the price per square meter dividing the price per month by the surface of the housing good. The main point of our study is to have geolocalized data. Consequently, we decide to drop observations which do not provide a city name or a precise geographic location.

Overall, our cleaning procedure decreases the number of observations by 12.87%. The largest part of this decrease is explained by observations which don't report the surface of the good. We believe that the price per square meter is the relevant statistics to characterize the housing market for several reasons. First, it provides a rental value which is used in other countries and is easily comparable. Second, it is used in the hedonic regres-

5. Enquêtes logements de l'INSEE, see http://www.insee.fr/fr/methodes/default.asp?page=definitions/enquete-logement.htm

6. Enquête Loyers et charges see http://www.insee.fr/fr/methodes/default.asp?page=sources/ope-enq-loyers-et-charges.htm

7. The French housing survey has 36 000 households but only 2 947 tenants in the private sector, the Survey on Rents and Housing Expenditures 4300 households.

8. http://www.observatoire-des-loyers.fr/

9. http://www.clameur.fr/

10. https://www.observatoires-des-loyers.org/2/accueil.htm

11. The 99th percentile of the price per square meter variable is 38.8 euros, the 99.9th percentile 63.5 euros, and the 99.95% 111.1 euros.

sion framework that we use in a second part of the study (see Musiedlak and Vignolles (2016))

This data set also contains a large number of variables and hedonic characteristics. Some variables as the surface, the number of rooms, the energy efficiency and green house gas emission or the precise location are directly embedded in the HTML code and easy to recover. Other variables as the type of contract or the amount of extra expenditures are recovered thanks to the description of the good using text mining algorithms. The most important variables and their creation procedure are summarized in the Appendix A.3.

4.3 The representativeness of the database

In order to assess the representativeness of the dataset coming from our collection process, we consider that housing units observed are a subsample of the exhaustive rental market which is observed in the French Census.

1. From the census, we create many strata crossing the location of the rental units (municipality) and their number of rooms. Each strata contains a number of observation in the census noted N^c

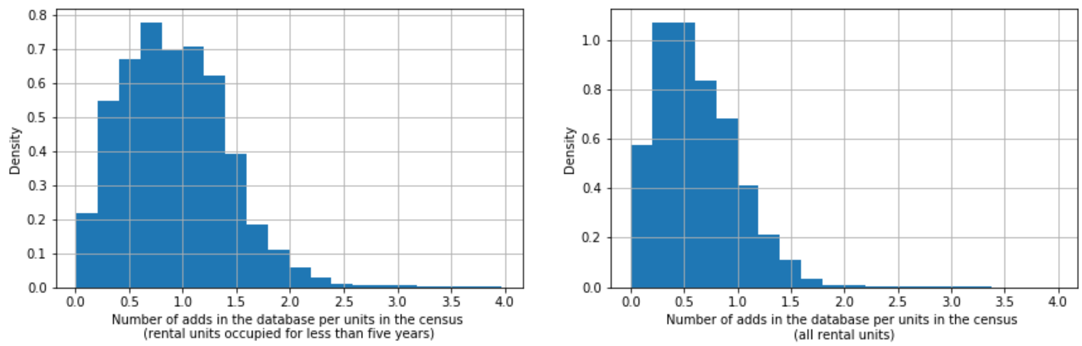
2. In a second step we assign our posted scraped to each strata. The number of scraped posts in each strata is noted n^s

3. The respresentativeness (i.e. number of post for each unit) is simply defined as $\frac{n^s}{N^c}$

We can thus measure the representativeness of each type of goods following two dimension: their location and the number of rooms. For example: we can know how many flats a post with two bedrooms in the Ist district of Paris will represent.

We use two different subsamples of the census to create two alternate measures. First, N^c is defined using all the rental units. Second, N^c is defined using the rental units occupied for less than five years used to proxy the flow of rental units on the market over our period of study. Figure 2 represents the distribution (weighted by the number of units) of the coverage of our strata. One can observe that On average each strata has 1 post per units rented for less than five years and 0.66 post per rental unit. Area with no coverage are rural municipalities with a residual rental sector. The coverage is usually very high in rural places where the number of tenants is low and in the suburbs while it is lower in city center where the number of tenants and the turnover in the private sector is very high. For example, within Paris the average coverage among strata is around 0.5 ads per rental units occupied for less than five years. This coverage is expected to grow over time as we only have been scraping for 2 years.

Figure 2: Representativeness of the database



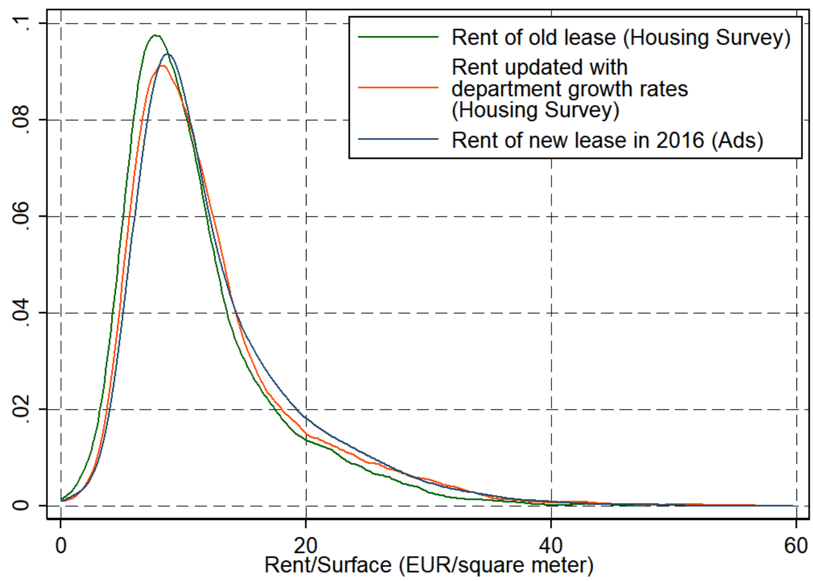
Each observation is a strata weighted by its number of rental units (N_c)

4.4 Is there a negotiation bias ?

If the type of housing unit observed and the channel used to find the flat appear fairly representative, one can fear that the posted rent might be different from the real one. Nevertheless several important observations lead us to believe that this bias remains limited. From a theoretical standpoint, if we model the housing market as a frictional market where a landlord and a tenant meet (Wheaton (1990)), the bargained rent is a weighted sum of the landlord's and the tenant's surpluses. The rent crucially depends on the relative bargaining power of the landlord / tenant. However, Desgranges and Wasmer (2000) show that when the bargaining power of the tenant is close to zero the rent converges toward the posted rent when we assume a price competition among landlords. Moreover, Binmore, Rubinstein, and Wolinsky (1986) show that the bargaining power in Nash bargaining process can be seen as a factor of relative impatience where the impatient party has a lower bargaining power. The lack of housing supply in France, particularly in large cities, leads us to believe that prospecting people have a relatively small bargaining process at least in the major urban areas. Moreover, for other markets, one can expect that the transparency of the the online ads where landlords can observe at a reduced cost the prices and movements of their competitors offering a similar unit in the same area can also drive the posted rent close to the market rent.

These reasons lead us to believe that posted rents are not likely to differ too much from the signed ones. To evaluate the importance of such a concern we confront our dataset with a standard survey based dataset : the French Housing Survey of 2013. The French housing survey is a series of survey representative at the national level performed every five years from a random sample of the French census. Its sample size is limited and contains information about 36 000 households and their housing conditions. It is only considered as representative for France, Ile de France (Paris region) and the North of France. Moreover, the number of tenants is relatively small as the sample only contains about 4 400 households in the private rental sector. In addition, the direct comparison with our dataset is limited as it only contains mostly old lease signed over the previous year. After deflating the rent with the Rent Revision Index in order to proxy for the signed rent, we update the rents of this survey with the corresponding departmental growth rate of new lease between the date the signature of the lease and 2016 published on the website CLAMEUR. As these growth rates are only published since 2000, this restricts the sample to 3 818 households. The change in the rent distribution is displayed in Figure 3.

Figure 3: Distribution of the rent per square meter



Source: Authors' computation from the French Housing Survey, updated with Clameur's department growth rates.

To assess the negotiation bias comparing both dataset, we estimate an hedonic regression model for each database in order to obtain an estimate of the rental value of a similar reference good for each region or department. We define this reference good as follows :

1. A flat with two rooms
2. with a surface of 50 square meters
3. located on the second floor
4. with no extra expenditures included in the rent

Formally, we follow a methodology close to Gouriéroux and Laferrère (2009), Musiedlak and Vignolles (2016) for housing price estimating the following model for the rent per square meter :

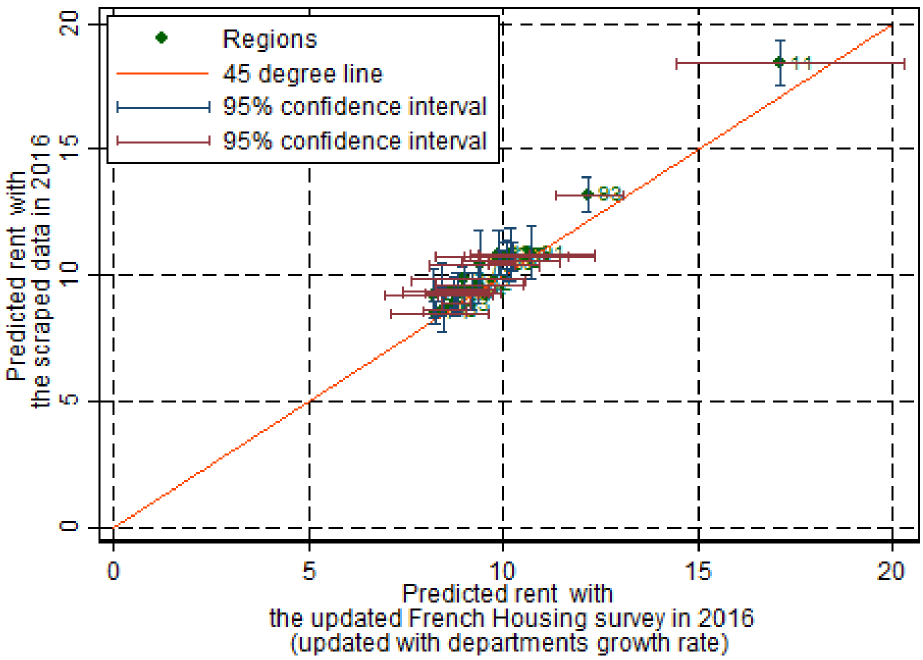
$$ln(r_{i,s}) = ln(r_s^{ref}) + X_{i,s}\beta + u_{i,s} \tag{9}$$

Centering the variables X_i around the reference good characteristics allows us to interpret the intercept $(ln(r_s^{ref}))$ as the log of the rent per square meter of the reference good in the department (or region) s. We control for the variables that are common to the two datasets: the log of the surface and its square, the number of rooms , the floor. The dependant variable is the log of the rent per square meter in the ads and the updated rent for the housing survey. Standard errors are clustered at the lowest geographical level available, departments for the French Housing Survey and Municipalities for the scraped Dataset.

Results of the hedonic regressions are reported in Table 10, one can notice that the coefficients between both database are fairly close in particular for the surface and its square. Moreover, when we confront our estimates of the rent per square meter of a representative good at the department or regional level with those based on the French housing survey, the correlation is higher than 90% in Figure 4 and 5. When plotting the 45 degree line, one does not observe any systematic bias in the market where the market is not tight; this should relieve our concerns about the negotiation bias. Even if the French Housing survey is only representative within Ile de France and the North of France, the strong correlation between these fixed effects can be considered as an important evidence of the limited bias in our scraped data when compared with standard surveys.Our estimates are only higher for Paris (department 75) and Hauts de Seine. However, our estimates for Paris are not very different from an alternate database dedicated to Paris area the OLAP. Here the average and median rent for new lease for a flat with 2 bedrooms in 2016 are respectively 24.8 and 25.1 euros per square meters for 1003 observations [12] which corresponds to our estimates for our reference good in Paris. Finally, a last remark arises. If a bias exists, it should be positive and our estimates of the cost of density would be an upper bound if we over estimate the difference in rents between small and large cities. This last consideration should relieve our concerns on the validity of our results if we find a lower elasticity than with housing prices.

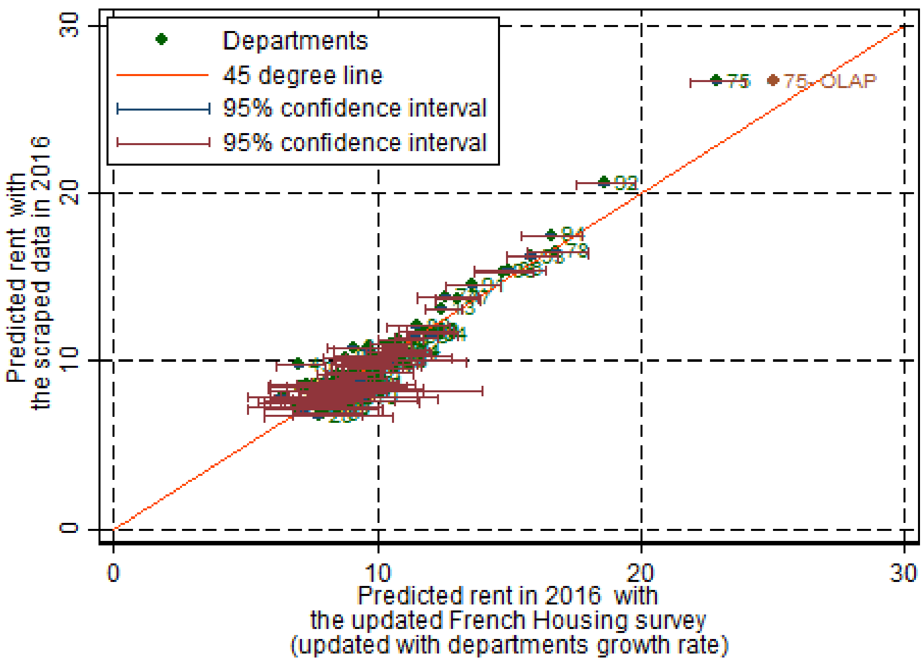
12. http://www.observatoire-des-loyers.fr/annees-precedentes/donnees-annee-2016

Figure 4: Predicted rent for a similar flat at the Regional level



Output of the column (1) and (2) in Table 3 resulting from the estimation of equation 9. Standard errors are clustered at the municipality level for the scraped Data and the Department level for the Housing Survey.

Figure 5: Predicted rent for a similar flat at the Department level



Output of the column (3) and (4) in Table 3 resulting from the estimation of equation 9. Standard errors are clustered at the municipality level for the scraped Data and the Department level for the Housing Survey.

4.5 Municipality characteristics and Urban Area controls

We also construct a dataset describing Municipalities and urban areas amenities as described in Combes, Duranton, and Gobillon (2018). As reported, we improve some of their instruments exploiting two original datasets. The first one is the dataset constructed in Motte, Séguy, and Théré (2003) which gathered the census for all French municipalities since 1793 and not only the largest one. This allows to recover the population of all urban areas since 1793. Second, we build a more precise measure of the temperature in January. We aslo gather data source on local taxation and compute the average tax rate per square meters using DGFIP and CEREMA files. Our sources are summarized in Table 11.

5 Results

The economic effects of density are a growing field of research. As cities are viewed as the result of agglomeration economies, economists have been investigating how density generates positive and negative externalities. In a recent paper, Ahlfeldt and Pietrostefani (2019) synthesise the net effects of density. They find that while density generates a large array of positive externalities as higher productivity, lower energy consumption, it goes along with large costs in particular for renters who face higher housing costs. They estimate that, if the net benefits of density are positive, the large elasticity of rents estimated in Combes, Duranton, and Gobillon (2018) might harm tenants. By measuring the relationship between density and rents instead of prices, we show that the cost for tenants might be lower and thus the net benefits of density much larger.

5.1 Comparing rents and rent price ratios across Urban Areas

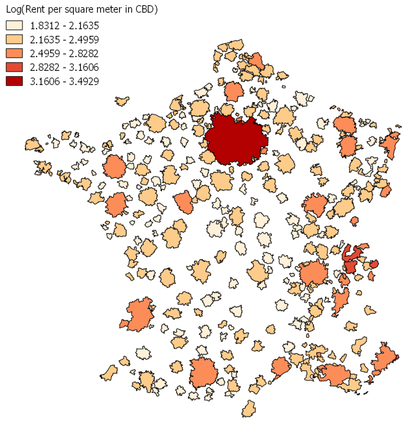
We first estimate Municipalities fixed effects thanks to equation 6. Our results for municipalities belonging to one urban area are summarized in Table 12 with the control variables we have at the municipal level. The log of the rent per square meter net of the hedonic characteristics range from 0.22 which is around 1.5 euros per square meters to 3.51 which is about 37 euros per square meters and correspond to the City center of Paris. Municipalities are on average 15 kilometers away from the barycenter of the Municipalities. Figure 11 represents the rent fixed effect for each municipalities as a function gradient of the distance from the City Business District in 4 large urban areas.

In the second step, we estimate the fixed effects of the urban area and their gradient thanks to equation 7. Panel a) in Figure 6 presents the fixed effects at the urban area level while the descriptive statistics for all the urban areas are summarized in Table 13. The urban area fixed effects (log(rent)) with and without control are extremely close and range from 1.82 to 3.66. This corresponds to a rent in the CBD going from 6.2 euros per square meters to 38.9 euros per square meter. Urban area fixed effects are also estimated without controls at the municipality levels in order to perform additional robustness checks as presented in Table 3 and 15. It is worth noting that these two steps can also be combined into one single step estimating directly urban area fixed effects from micro data, results remain qualitatively unchanged. We also confront these rent fixed effects with the price fixed effects estimated in Combes, Duranton, and Gobillon (2018). We approximate the rent price ratio as follows $R/P = e^{C^r}/(e^{C^P} \times 10000)$, this is only an approximation as the estimated fixed effects should be corrected with the error terms and

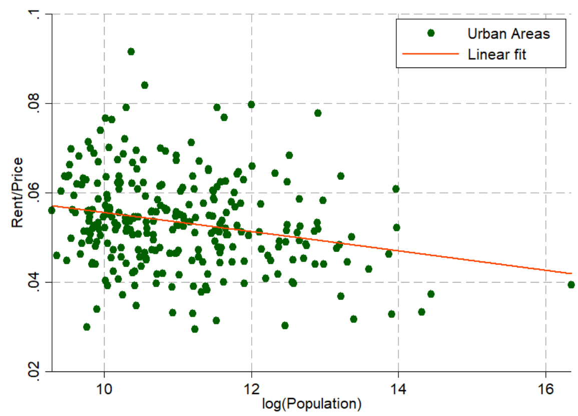
Urban Area fixed effects are provided in relative terms with respect to Paris in Combes, Duranton, and Gobillon (2018) while we don't have the intercept. We take 10,000, as the value of one square meter in the CBD of Paris in 2017. As expected from equation 3, the resulting Rent Price Ratios for urban areas are displayed in Figure 6 and appear negatively correlated with city size and price growth and positively correlated with the tax rate. The rent and price gradients appear relatively close when confronting our results with the gradients estimated in Combes, Duranton, and Gobillon (2018).

In Table 14 in the Appendix, we rewrite equation 3 as the rent price ratio and estimate correlations between the rent price ratio, city size and the user cost parameters available. As expected, the inclusion of tax and price growth affects the relationship between city size and the rent price ratio. It is worth noting that this finding is in line with Gyourko, Mayer, and Sinai (2013) who emphasizes that the rent price ratio in large and superstar cities is usually lower as people expect rents to increase faster in these very productive areas. More importantly, this confirms the idea that not accounting for the parameters of the user cost might lead to overestimate the elasticity of housing costs with respect to city population when estimated with housing prices. Moreover, one can note that even controlling for these parameters, the negative correlation between density and the rent price ratio persists, there might be other unobserved factor as interests rates likely to affect this relationship. Our proxy for house price growth based on past growth is probably not fully satisfactory neither. In our view, this advocates for an estimate of the cost of density based on rents rather than price when possible in order to avoid to compute a user cost which might rely on strong assumptions.

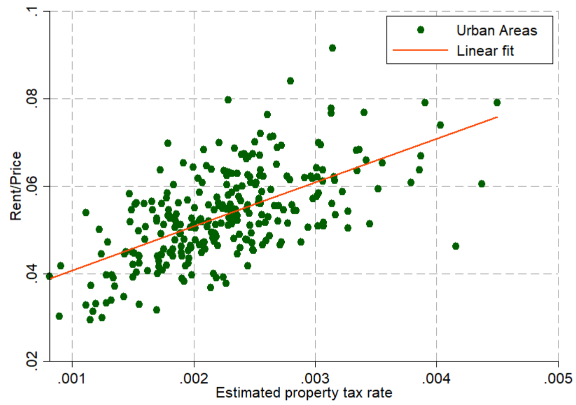
Figure 6: Rents and Rent price ratios in the center of French Urban Areas



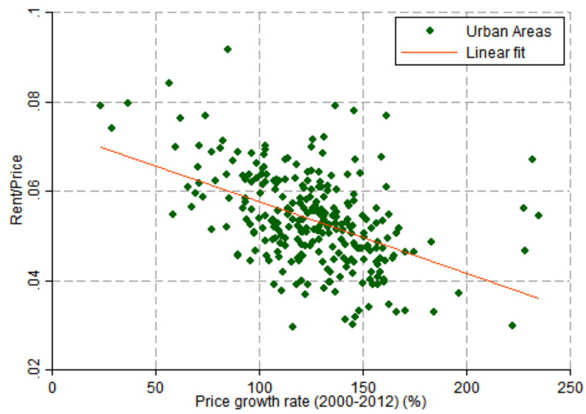
(a) Log of the rent per square meter



(b) Rent Price ratio and city population



(c) rent Price ratio and property tax



(d) Rent Price ratio and price growth

5.2 Estimates of the elasticity of rents with respect to density

Finally, we estimate the elasticity of housing cost with respect to city population. We use the Urban areas fixed effects estimated in the previous section combined with two data sets containing the characteristics and amenities for all French Urban Areas in mainland France. The first one corresponds to the original data set built in Combes, Duranton, and Gobillon (2018) while the second builds and update this data set from their instructions and new data sources as explained above. We estimate equation 8 and report our main OLS results in Table 3. Provided the relatively small changes involved by the Instrumental Variable Strategy we treat OLS as our favourite specification in line with the previous study on prices. Panel A) reproduces Combes, Duranton, and Gobillon (2018) main specification with their data set and code. Results are the same as the dependent variable is still housing prices. Panel B) introduces our rent fixed effects in their data set as a dependent variable. The resulting elasticity is significantly lower than with prices. As our rents were measured in 2016, panel C) updates the control and dependant variables with a self constructed datasets detailed in Table 11. Our results remain virtually unchanged and still lower than when estimated with prices. Finally, as Motte, Séguy, and Théré (2003) allows us to compute their population instrument for all the urban areas in France, we also increase the estimated sample which confirms the lower magnitude of the cost of agglomeration. The residuals and predicted relationship between population and rent in the center are represented in Figure 12. As this introduces two outliers which are very touristic areas in the Gulf of Saint Tropez, for presentation purpose only, and not in the main estimates we add a dummy to control for the specificity of these two urban areas. As a final robustness check, we reproduce our estimates instrumenting the population and land area in Table 15 in the appendix. Results remain qualitatively unchanged, 2SLS estimates are always lower with rents than with prices and suggest that the elasticity could even be slightly lower. If we follow the authors' favourite specification, an elasticity of 0.125 estimated on our full sample should be retained which results in a cost of agglomeration which range between 0.022 for a city of 100 000 inhabitants (i.e 2/3 of the one estimated with prices) to 0.049 for a city as Paris (instead of 0.08112 with prices). In the 2SLS estimates the cost of agglomeration could even be lower as the elasticity oscillates around 0.08 with the extended sample which would result in a cost of agglomeration of 0.014 for a city of 100 000 inhabitants which would be half of the one estimated with prices. However, as this result only relies on the extended sample where many cities are extremely small, we keep the elasticity of 0.125 as our favourite parameter.

Table 3: The determinants of housing rent at the city center, OLS regression

	(1)	(2)	(3)	(4)	(5)	(6)
First-step	Only Fixed Effects			Full set of controls		
Controls	N	Y	Ext.	N	Y	Ext.
Panel A: Original specification using price fixed effects						
Log(Population)	0.217***	0.176***	0.224***	0.252***	0.208***	0.304***
	(0.0160)	(0.0111)	(0.0223)	(0.0198)	(0.0137)	(0.0284)
Log(Land area)	-0.151***	-0.153***	-0.224***	-0.143***	-0.152***	-0.276***
	(0.0168)	(0.0107)	(0.0231)	(0.0185)	(0.0135)	(0.0294)
N	1937	1937	1937	1937	1937	1937
R^2	0.352	0.645	0.720	0.403	0.659	0.726
Panel B: substituting price fixed effect with rent fixed effect						
Log(Population)	0.177***	0.146***	0.193***	0.170***	0.147***	0.223***
	(0.0164)	(0.0132)	(0.0245)	(0.0152)	(0.0137)	(0.0250)
Log(Land area)	-0.0314**	-0.0419***	-0.104***	-0.0719***	-0.0818***	-0.171***
	(0.0135)	(0.0112)	(0.0231)	(0.0133)	(0.0122)	(0.0234)
N	277	277	277	277	277	277
R^2	0.625	0.750	0.790	0.544	0.668	0.748
Panel C: and updating population and control variables						
Log(Population)	0.172***	0.138***	0.0920***	0.167***	0.138***	0.101***
	(0.0124)	(0.0128)	(0.0290)	(0.0117)	(0.0126)	(0.0285)
Log(Land area)	-0.0289**	-0.0385***	-0.0133	-0.0713***	-0.0787***	-0.0586**
	(0.0145)	(0.0133)	(0.0287)	(0.0137)	(0.0130)	(0.0282)
N	277	277	277	277	277	277
R^2	0.619	0.709	0.783	0.541	0.621	0.717
Panel D: and extending the sample						
Log(Population)	0.169***	0.131***	0.109***	0.159***	0.125***	0.117***
	(0.0135)	(0.0125)	(0.0213)	(0.0129)	(0.0127)	(0.0214)
Log(Land Area)	-0.0511***	-0.0527***	-0.0363^c	-0.0817***	-0.0837***	-0.0801***
	(0.0137)	(0.0116)	(0.0210)	(0.0132)	(0.0117)	(0.0211)
N	352	352	352	352	352	352
R^2	0.446	0.640	0.778	0.355	0.535	0.719

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Notes: The dependent variable is an urban area fixed effect estimated in the previous section for panel b), c) and d) or in Combes, Duranton, and Gobillon (2018) for panel a). Standard errors clustered at the urban area level are between brackets. For second-step controls, N, Y, and Ext. stand for no further explanatory variables beyond population, land area, and year effects, a set of explanatory variables, and a full set, respectively. Second-step controls include population growth of the urban area (as log of 1 + annualised population growth over the period), income and education variables for the urban area (log mean income, log standard deviation (with the exception of panel C and D) , and share of university degrees). Extended controls additionally include the urban-area means of the same 20 geography and geology controls and the same two land use variables (share of built-up land and average height of buildings)

This discrepancy between our results with rents and the original estimates with price are non negligible and of significant economic importance. They might be driven by the fact that housing prices can remain higher than their fundamental values for relatively long period of times as documented in Ambrose, Eichholtz, and Lindenthal (2013). Alternatively, this can also reflect the fact that denser areas also have higher expected growth rate resulting in an increase in housing prices as emphasized in Gyourko, Mayer, and Sinai (2013). Our estimates are of particular importance to understand the benefits of density and to en-light the debate on compact cities (Glaeser (2011)). Indeed, evidence so far suggests that the positive externalities on income appear lower than the increase in housing costs (Ahlfeldt and Pietrostefani (2019)). However Table 4 reproduces the accounting exercise performed by the authors introducing the implication of our updated estimates of the impact of density on rents.

Table 4: Updated Monetised effects of a 10% increase in density (Ahlfeldt and Pietrostefani (2019))

ID	Outcome Category	Factor Income	Quality of Life	Amenity value	Effect on Owner	Renter	External Welfare
1	Wage	140	-71	0	71	71	0
2	Innovation	0	0	0	0	0	2
3	Value of space	**144** (243)	**144** (243)	0	0	**-144** (-243)	0
4	Job accessibility	0	0	62	62	62	0
5	Service access	0	0	0	49	49	49
6	Eff; of pub. serv deli.	0	0	0	0	0	21
7	Social equity	0	0	0	0	0	-6
8	Safety	0	0	8	8	8	0
9	Urban green	0	0	41	41	41	0
10	Pollution reduction	0	0	14	14	14	0
11	Energy efficiency	0	0	25	25	25	0
12	Traffic flow	0	0	-35	-35	-35	0
13	Sustainable mode choice	0	0	0	0	0	0
14	Health	0	0	-32	-32	-32	0
15	Subj well being	0	0	-26	-26	-26	0
	Sum	383	**70** (172)	106	177	**33** (-64)	29

Note: Accounting exercise borrowed from Ahlfeldt and Pietrostefani (2019). **Bold values are the updated values with our estimates based on rental data**, values between brackets are the original estimates based on Combes, Duranton, and Gobillon (2018) using prices.
Reading: In an average city in a high income country, a rise of 10% of the density would increase the wage of a renter by 71 dollars, its rent by 144 dollars (instead of 243 when estimated with price). The net total benefits of density would be of 33 dollars per year (instead of -64).

This exercise assesses the monetary costs and benefits of density for a person living in an average urban area in a high income country. If this person earns 35,000 dollars per year, a 10% increase in density would result in an increase of 140 dollars of its wage and a rise of 243 dollars of housing price. Nevertheless, our estimates suggest that the increase in rent could be limited to 144. As a result, one can observe that the net benefits of density are now also positive for tenants. This would help rationalizing why tenants are also more represented in denser places. Moreover, this order of magnitude also suggests that the two main benefits and costs of agglomeration tend to be of the exact same order of magnitude while the additional positive externalities as the increase in energy efficiency and the reduction in traffic flows would justify the public support for compact cities. To have a clear interpretation of our results, one should bear in mind that this elasticity corresponds to a long run relationship which might be affected by the stringency of the housing supply. As documented in Saiz (2010), Severen and Plantinga (2018), and Hilber and Vermeulen (2015), land use regulation as height restriction or zoning and geographical obstacles might reduce the supply elasticity and associate city growth with higher housing costs on shorter time horizon (Combes, Duranton, and Gobillon (2018)).

6 Conclusion

In this paper, we argue that rents should be preferred to prices in order to measure the costs of agglomeration. Indeed, the cost of housing of homeowners depends on several parameters as interest rates or taxes that might vary between urban areas. As a consequence, an upper bound of the cost of agglomeration can be proxied with market rent as buyers should at least be indifferent between owning and renting. If access to rental data can sometimes appear more difficult, we describe a data collection based on webscraping that can allow to gather reliable information on the rental market. If these online data correspond to posted rents and not to signed contracts, one can think that the relative transparency of online platforms tends to force landlords to reveal the market price. The comparison between our data set and standard surveys as the French housing survey supports this intuition. Indeed, one can observe that the predicted rents using both data set are extremely close.

When estimating the elasticity between housing cost and the size of the agglomeration thanks to our rental dataset, the estimated cost of agglomeration is almost divided by 2 for a large agglomeration as Paris when compared with Combes, Duranton, and Gobillon (2018). Our estimates bring new data to the debate trying to assess the opportunity of compact cities (Ahlfeldt and Pietrostefani (2019)). Showing that the cost of agglomeration might be lower than what was previously estimated, our results suggest that increasing density might also be beneficial to tenants.

As a conclusion, we would like to emphasize that the discrepancy between prices and rents might also be a result per se that shouldn't be neglected and would deserve additional research. As prices are forward looking, one can expect that the larger agglomeration cost estimated with price could reflect that denser place might also have higher expected growth. The success of superstar cities showing sustained income growth rate (Gyourko, Mayer, and Sinai (2013)) might be related with their density.

References

Ahlfeldt, Gabriel M, and Elisabetta Pietrostefani. 2019. "The economic effects of density: A synthesis."

Albouy, David Y. 2008. *Are Big Cities Really Bad Places to Live?: Improving Quality-of-life Estimates Across Cities.* 14472. Citeseer.

Albouy, David, Gabriel Ehrlich, and Minchul Shin. 2018. "Metropolitan land values." *Review of Economics and Statistics* 100 (3): 454–466.

Ambrose, Brent W, Piet Eichholtz, and Thies Lindenthal. 2013. "House prices and fundamentals: 355 years of evidence." *Journal of Money, Credit and Banking* 45 (2-3): 477–491.

Arnott, Richard. 1995. "Time for revisionism on rent control?" *The Journal of Economic Perspectives* 9 (1): 99–120.

Basten, Christoph, Maximilian Von Ehrlich, and Andrea Lassmann. 2017. "Income Taxes, Sorting and the Costs of Housing: Evidence from Municipal Boundaries in Switzerland." *The economic journal* 127 (601): 653–687.

Binmore, Ken, Ariel Rubinstein, and Asher Wolinsky. 1986. "The Nash bargaining solution in economic modelling." *The RAND Journal of Economics:* 176–188.

Boeing, Geoff, and Paul Waddell. 2016. "New Insights into Rental Housing Markets across the United States: Web Scraping and Analyzing Craigslist Rental Listings." *Journal of Planning Education and Research:* 0739456X16664789.

Bonnet, Odran, Pierre-Henri Bono, Guillaume Chapelle, Etienne Wasmer, et al. 2014. "Does housing capital contribute to inequality? A comment on Thomas Piketty's Capital in the 21st Century." *Sciences Po Economics Discussion Papers* 7:12.

Bricongne, Jean-Charles, Peter Pontuch, and Alessandro Turrini. 2017. "Insights from a new dataset of house prices in level." *European Commission, DG ECFIN.*

Brülhart, Marius, Jayson Danton, Raphaël Parchet, and Jörg Schläpfer. 2017. "Who Bears the Burden of Local Income Taxes." *HEC Lausanne mimeo.*

Campbell, Sean D, Morris A Davis, Joshua Gallin, and Robert F Martin. 2009. "What moves housing markets: A variance decomposition of the rent–price ratio." *Journal of Urban Economics* 66 (2): 90–102.

Chapelle, Guillaume, Etienne Wasmer, and Pierre-Henri Bono. 2017. "Reducing transportation costs in an urban labor market with frictional housing markets." *Sciences Po mimeo.*

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2018. "The Costs of Agglomeration: House and Land Prices in French Cities" (October). doi:`10.1093/restud/rdy063`. `https://dx.doi.org/10.1093/restud/rdy063`.

Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The empirics of agglomeration economies." In *Handbook of regional and urban economics,* 5:247–348. Elsevier.

Davis, Donald R, Jonathan I Dingel, Joan Monras, and Eduardo Morales. 2017. *How Segregated is Urban Consumption?* Technical report. National Bureau of Economic Research.

Davis, Morris A, and Jonathan Heathcote. 2007. "The price and quantity of residential land in the United States." *Journal of Monetary Economics* 54 (8): 2595–2620.

Desgranges, Gabriel, and Etienne Wasmer. 2000. "Appariements sur le Marché du Logement." *Annales d'Economie et de Statistique:* 253–287.

DiPasquale, Denise, and William C Wheaton. 1996. *Urban economics and real estate markets.* Vol. 23. 7. Prentice Hall Englewood Cliffs, NJ.

Edelman, Benjamin G, and Michael Luca. 2014. "Digital discrimination: The case of Airbnb. com."

Fujita, M, and J-F Thisse. 2002. *Economics of Aglomeration: Cities, Industrial Location, and Regional Growth.*

Gallin, Joshua. 2008. "The long-run relationship between house prices and rents." *Real Estate Economics* 36 (4): 635–658.

Garner, Thesia I, and Randal Verbrugge. 2009. "Reconciling user costs and rental equivalence: Evidence from the US consumer expenditure survey." *Journal of Housing Economics* 18 (3): 172–192.

Glaeser, Edward. 2011. *Triumph of the City.* Pan.

Glaeser, Edward L, and Joseph Gyourko. 2007. *Arbitrage in housing markets.* Technical report. National Bureau of Economic Research.

Glaeser, Edward L, and Charles G Nathanson. 2015. "Housing bubbles." In *Handbook of regional and urban economics,* 5:701–751. Elsevier.

———. 2017. "An extrapolative model of house price dynamics." *Journal of Financial Economics* 126 (1): 147–170.

Gouriéroux, Christian, and Anne Laferrère. 2009. "Managing hedonic housing price indexes: The French experience." *Journal of Housing Economics* 18 (3): 206–213.

Gyourko, Joseph, Christopher Mayer, and Todd Sinai. 2013. "Superstar cities." *American Economic Journal: Economic Policy* 5 (4): 167–99.

Halket, Jonathan, Lars Nesheim, and Florian Oswald. 2015. *The housing stock, housing prices, and user costs: the roles of location, structure and unobserved quality.* Technical report. cemmap working paper, Centre for Microdata Methods and Practice.

Hijmans, Robert J, Susan E Cameron, Juan L Parra, Peter G Jones, and Andy Jarvis. 2005. "Very high resolution interpolated climate surfaces for global land areas." *International journal of climatology* 25 (15): 1965–1978.

Hilber, Christian AL, and Wouter Vermeulen. 2015. "The impact of supply constraints on house prices in England." *The Economic Journal* 126 (591): 358–405.

Hill, Robert J, and Iqbal A Syed. 2016. "Hedonic price–rent ratios, user cost, and departures from equilibrium in the housing market." *Regional Science and Urban Economics* 56:60–72.

Himmelberg, Charles, Christopher Mayer, and Todd Sinai. 2005. "Assessing High House Prices: Bubbles, Fundamentals and Misperceptions." *Journal of Economic Perspectives* 19 (4): 67–92.

Hyland, Marie, Ronan C Lyons, and Sean Lyons. 2013. "The value of domestic building energy efficiency—evidence from Ireland." *Energy Economics* 40:943–952.

Jiang, Bin, and Jean-Claude Thill. 2015. *Volunteered Geographic Information: Towards the establishment of a new paradigm.*

Kholodilin, Konstantin A, Andreas Mense, and Claus Michelsen. 2017. "The market value of energy efficiency in buildings and the mode of tenure." *Urban Studies* 54 (14): 3218–3238.

Laouénan, Morgane, and Roland Rathelot. 2017. "Ethnic Discrimination on an Online Marketplace of Vacation Rental."

Loberto, Michele, Andrea Luciani, and Marco Pangallo. 2018. "The potential of big housing data: an application to the Italian real-estate market."

Mense, Andreas, Claus Michelsen, and Konstantin A Cholodilin. 2017. *Empirics on the causal effects of rent control in Germany.* Technical report. FAU Discussion Papers in Economics.

Motte, Claude, Isabelle Séguy, and Christine Théré. 2003. *Communes d'hier, communes d'aujourd'hui: les communes de la France métropolitaine, 1801-2001: dictionnaire d'histoire administrative.* Ined.

Musiedlak, Yoann, and Benjamin Vignolles. 2016. "Les mouvements des prix immobiliers dans l 'ancien au cours des années 2000 : des marchés locaux différenciés." *Document de travail du CGEDD,* no. 24: 1–40.

Poterba, James M. 1984. "Tax subsidies to owner-occupied housing: an asset-market approach." *The quarterly journal of economics* 99 (4): 729–752.

Saiz, Albert. 2010. "The geographic determinants of housing supply." *The Quarterly Journal of Economics* 125 (3): 1253–1296.

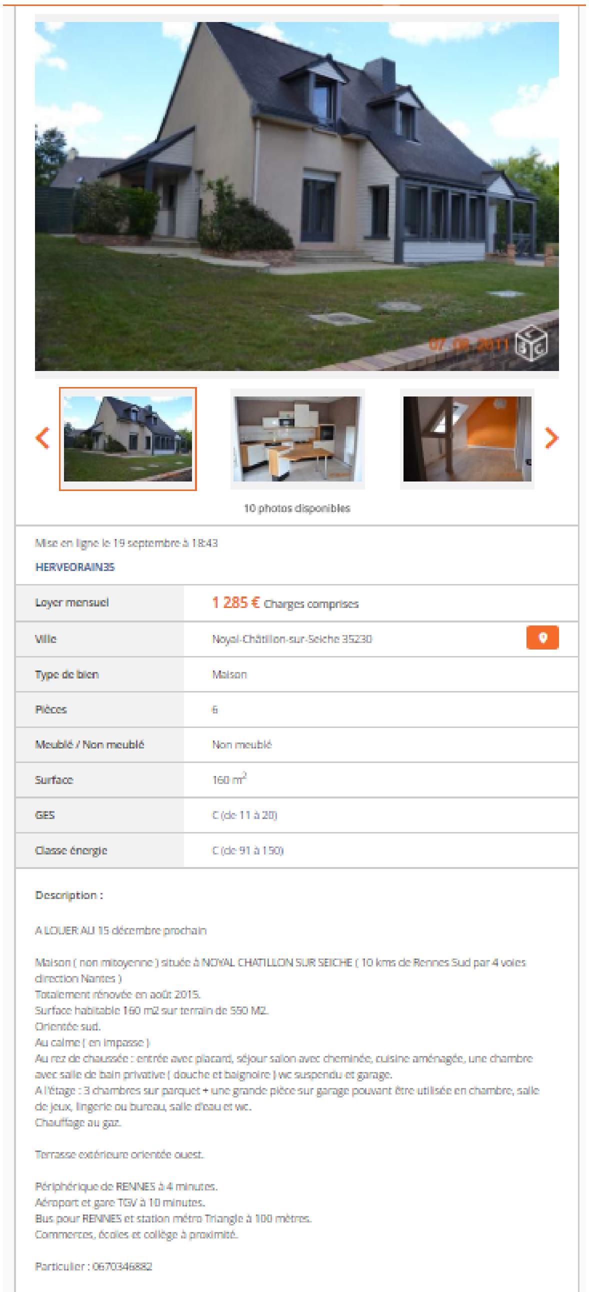
Severen, Christopher, and Andrew J Plantinga. 2018. "Land-use regulations, property values, and rents: Decomposing the effects of the California Coastal Act." *Journal of Urban Economics* 107:65–78.

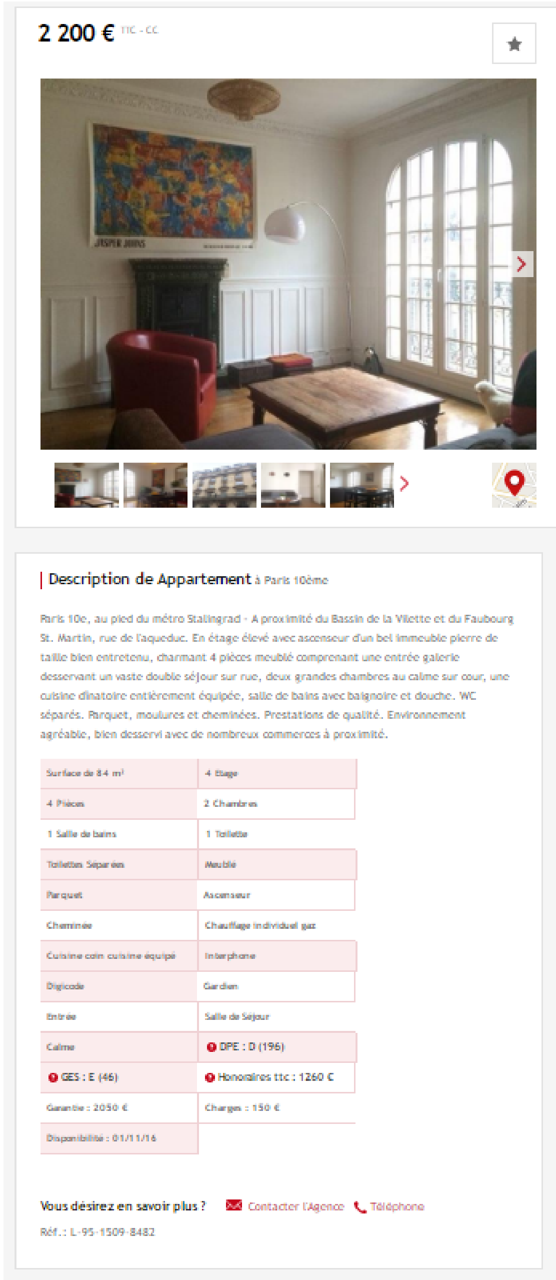
Verbrugge, Randal. 2008. "The puzzling divergence of rents and user costs, 1980–2004." *Review of Income and Wealth* 54 (4): 671–699.

Wheaton, William C. 1990. "Vacancy, search, and prices in a housing market matching model." *Journal of Political Economy:* 1270–1292.

A Rental Dataset

A.1 Example of Posts and Location





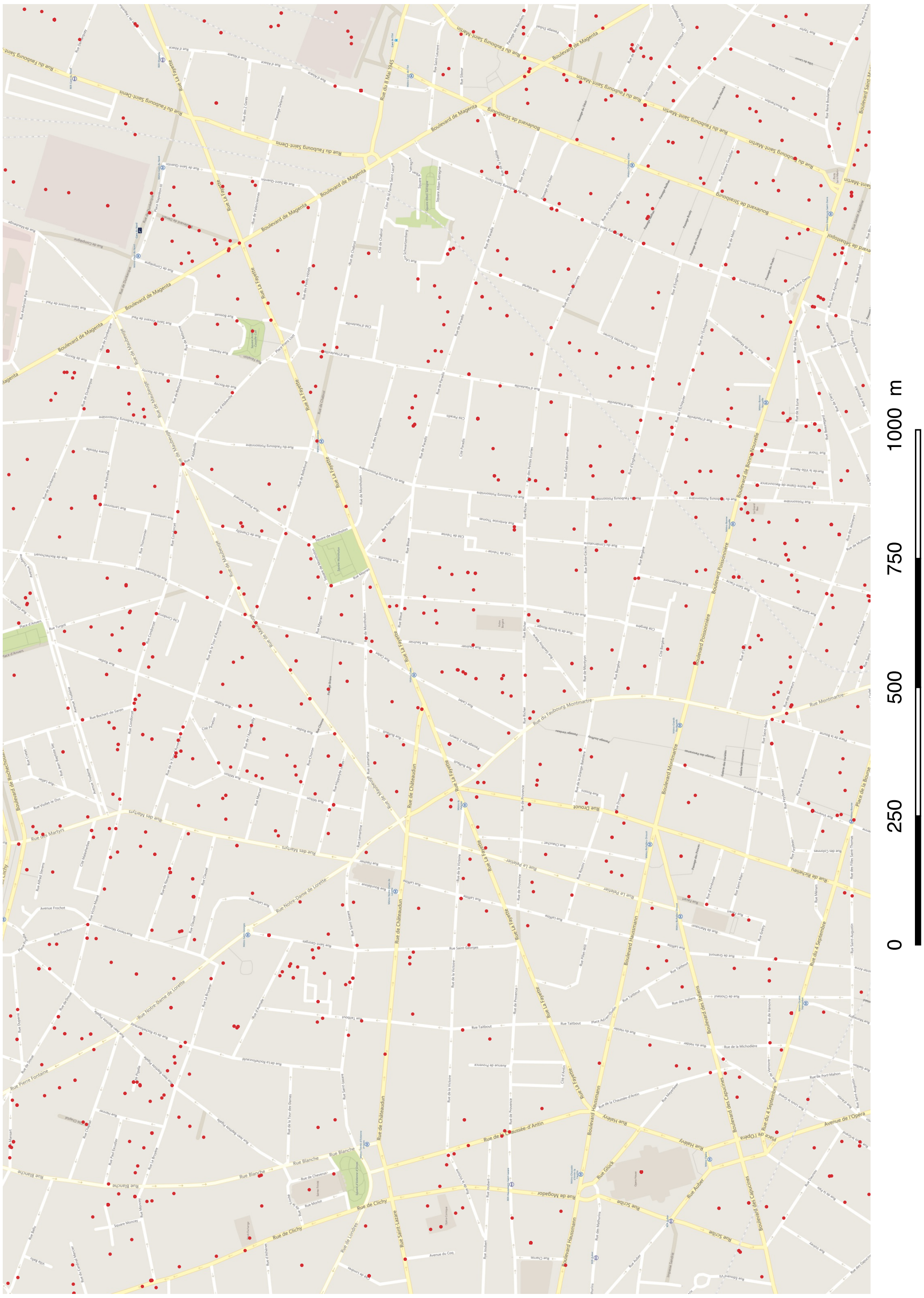


Figure 7: Posts location around Grands Boulevards area in Paris

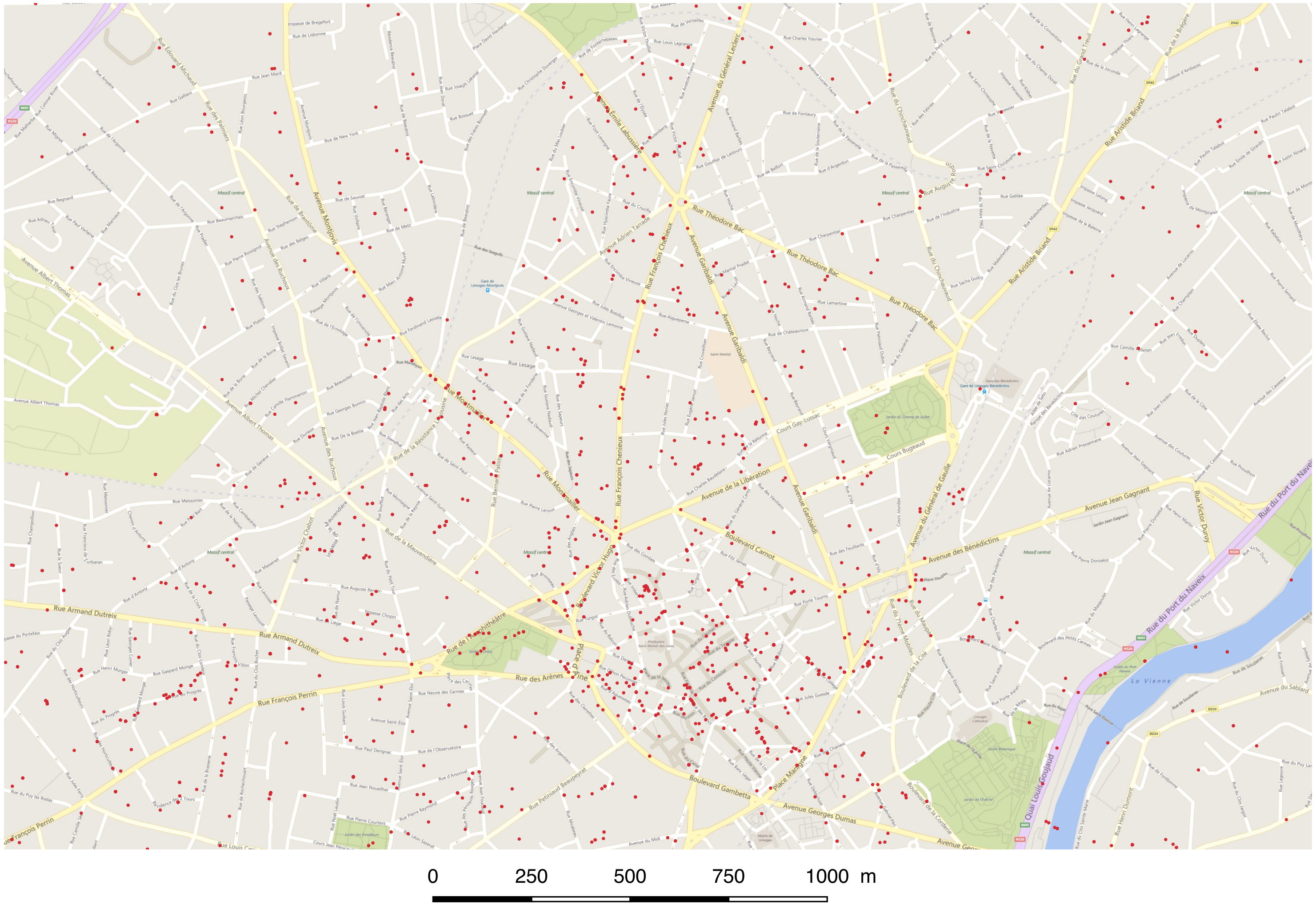
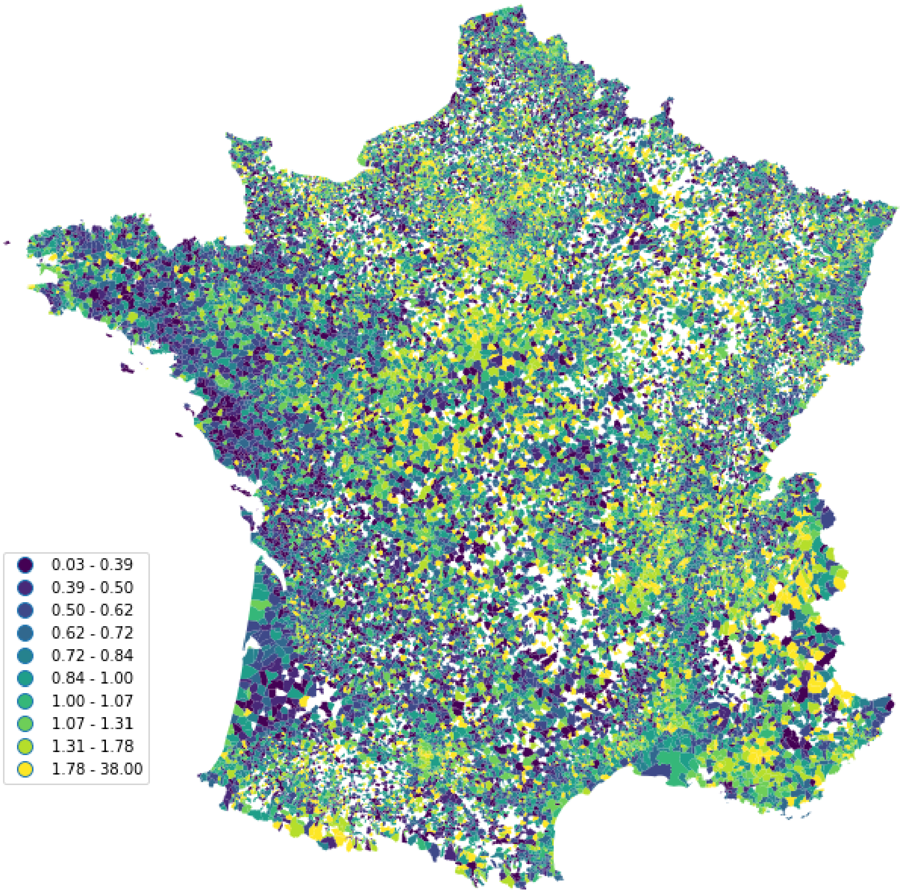


Figure 8: Posts location for Limoges

A.2 Representativeness

Figure 9: Representativeness of the database through space



This represents the average number of ads per unit in all strata of a Municipality

A.3 Variables

A.3.1 The geocoded location and the granularity of the dataset

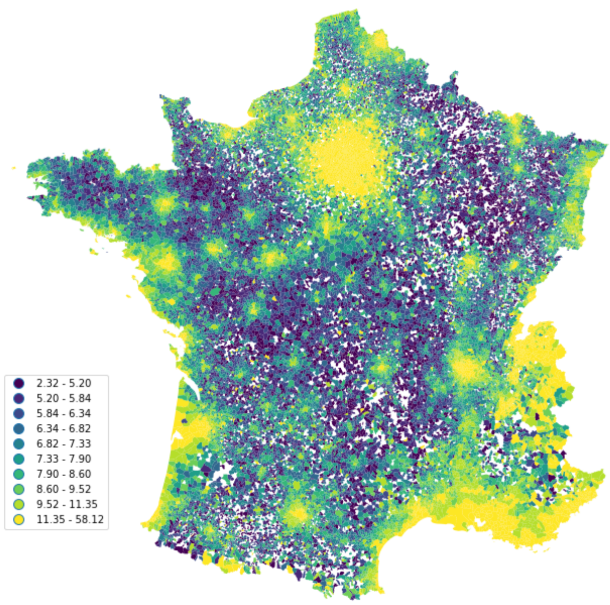
One important variable is the location of each post. Both websites provide geocoded information for each good. However, some realtors or households might not be willing to disclose too precisely the address even if platforms usually provide financial incentives to disclose the true location of the good. The HTML code informs directly to what level of precision the geolocation corresponds. Table 5 summarizes the level of geocoding in our database. 60% of the ads are located at the broadest level : French Municipalities while 40% remaining are precisely geocoded at the address or neighborhood level using the information provided by the user or the location of the device used by the user when creating the add. This database provides thus fine grain data as even municipalities remain quite small.

This allows us to compute an average rent for the majority of the municipalities in France as illustrated in Figure 10 where one can easily identify the main urban areas and the places close to the frontiers where rents are usually higher.

Table 5: Precision of the geocoding

Geocoding (%) Unknown:	0.0
Geocoding (%): address	17.9
Geocoding (%): browser	0.0
Geocoding (%): city	59.0
Geocoding (%): device	0.2
Geocoding (%): neighborhood	12.8
Geocoding (%): user	10.0
Geocoding (%): zipcode	0.0

Figure 10: Average rent in French Municipalities



Gross average rent per square meter of online ads

A.3.2 The type of units, the surface and the number of rooms

Each website has a specific part of the webpage dedicated to the type of unit; the surface and the number of rooms. No treatment is thus required and these variables are taken directly from the HTML code.

Table 6: Type of units , number of rooms and surface

	count	mean	std	min	25%	50%	75%	max
surface	4225940.0	55.9	31.1	1.0	34.0	50.0	70.0	1080.0
single unit (%)	4225940.0	15.8	36.4	0.0	0.0	0.0	0.0	100.0
Rooms (%): 01	4225940.0	20.7	40.5	0.0	0.0	0.0	0.0	100.0
Rooms (%): 02	4225940.0	32.5	46.9	0.0	0.0	0.0	100.0	100.0
Rooms (%): 03	4225940.0	26.2	44.0	0.0	0.0	0.0	100.0	100.0
Rooms (%): 04	4225940.0	12.6	33.2	0.0	0.0	0.0	0.0	100.0
Rooms (%): 05	4225940.0	5.5	22.8	0.0	0.0	0.0	0.0	100.0
Rooms (%): 6+	4225940.0	2.5	15.7	0.0	0.0	0.0	0.0	100.0

As one can observe in Table 6 , most of the units are flats as single units only represent about 16% of the sample. Units are of a relatively small size as their vast majority have one or two rooms while the average surface is about 56 square meters. These characteristics are typical from the French rental market that is dedicated to younger people with few children.

A.3.3 The rent, type of lease and the expenditures

Another important variable is the rent. This variable is also directly coded and easy to recover. The average gross rent is about 650 euros while the rent per square meter is around 13 euros. Both websites also provide additional information specifying whether the rent displayed includes extra expenditures (as waste collection, water , heating). 70% of the rent displayed includes some kind of extra expenditures. Unfortunately, the share of the rent attributed to these is not directly coded and is recovered from the text using regular expressions. The algorithm identifies whether the word ”charges” is in the text and recover the amount in euro around this word that is inferior to the rent. About 30% of the ads inform the amount of extra expenditures. The average estimated amount of extra expenditures on the subsample is around 58 euros which represents 9% of the average rent. From the text it is also possible to infer which type of expenditures are included as collective heating or trash collection. Finally, a second important information is the type of lease indicating whether furnitures are included in the lease or not. This variable is of particular importance as the minimal length of the lease is 1 year when the flat is furnished while it will be 3 years when not. Once again, if this information appears in the code of the web page for the most recent period, this was not systematically filled in the first waves. Consequently it is also coded from regular expressions identified in the description. About 20% of the flats are offered as furnished.

Table 7: Price, expenditures and type of lease

	count	mean	std	min	25%	50%	75%	max
Rent	4225940.0	646.2	396.7	8.0	445.0	561.0	730.0	65000.0
Rent per square meter	4225940.0	13.3	6.7	2.0	8.9	11.7	16.0	100.0
Expenditures : Included	4225940.0	72.3	44.8	0.0	0.0	100.0	100.0	100.0
Expenditures : Not Included	4225940.0	5.9	23.6	0.0	0.0	0.0	0.0	100.0
Expenditures : Unknown	4225940.0	21.8	41.3	0.0	0.0	0.0	0.0	100.0
Expenditures	1594691.0	58.3	51.9	0.0	30.0	45.0	72.0	3705.0
Collective heating (%)	4225940.0	3.5	18.4	0.0	0.0	0.0	0.0	100.0
Hot water (%)	4225940.0	0.2	4.1	0.0	0.0	0.0	0.0	100.0
Trash collection (%)	4225940.0	4.7	21.1	0.0	0.0	0.0	0.0	100.0
Furnished (%): No	4225940.0	81.1	39.2	0.0	100.0	100.0	100.0	100.0
Furnished (%): Yes	4225940.0	18.9	39.2	0.0	0.0	0.0	0.0	100.0

A.3.4 Floors and other amenities

It is also possible to identify in the description what is the floor and the amenities of the building. As one can see in Table 8, the floor can be recovered for 40% of the ads while 14% of the ads announce the presence of an elevator. 36% have a balcony or a kitchen with some equipment. Finally, 46% offer some possibilities to park a car.

Table 8: Floors and other amenities

	count	mean	std	min	25%	50%	75%	max
Floor								
Floor (%): 0.0	4225940.0	8.9	28.5	0.0	0.0	0.0	0.0	100.0
Floor (%): 1.0	4225940.0	10.8	31.1	0.0	0.0	0.0	0.0	100.0
Floor (%): 2.0	4225940.0	7.9	27.0	0.0	0.0	0.0	0.0	100.0
Floor (%): 3.0	4225940.0	4.1	19.9	0.0	0.0	0.0	0.0	100.0
Floor (%): 4.0	4225940.0	2.0	13.9	0.0	0.0	0.0	0.0	100.0
Floor (%): 5.0	4225940.0	0.9	9.7	0.0	0.0	0.0	0.0	100.0
Floor (%): 6+	4225940.0	1.0	10.0	0.0	0.0	0.0	0.0	100.0
Floor (%): Unknown floor	4225940.0	59.6	49.1	0.0	0.0	100.0	100.0	100.0
Floor (%): last floor	4225940.0	4.7	21.1	0.0	0.0	0.0	0.0	100.0
Amenities								
Elevator (%)	4225940.0	14.0	34.7	0.0	0.0	0.0	0.0	100.0
Double glazing (%)	4225940.0	9.5	29.3	0.0	0.0	0.0	0.0	100.0
Kitchen with equipment (%)	4225940.0	35.5	47.9	0.0	0.0	0.0	100.0	100.0
garage (%)	4225940.0	46.1	49.8	0.0	0.0	0.0	100.0	100.0
Garden (%)	4225940.0	17.4	37.9	0.0	0.0	0.0	0.0	100.0
Balcony (%)	4225940.0	35.6	47.9	0.0	0.0	0.0	100.0	100.0

A.3.5 Energy consumption and Greenhouse Gas emission

Since July 2007, each landlord should realize a diagnosis of the energy efficiency to rent their unit. This information was already used in previous work in order to investigate the impact of energy efficiency of buildings on real estate prices (Kholodilin, Mense, and Michelsen (2017) and Hyland, Lyons, and Lyons (2013)). These previous work emphasize the importance of these indicators showing a positive correlation between rent, prices and the energy efficiency displayed. The energy efficiency and GES consumption category are directly embedded inside the HTML code and can be easily recovered. As illustrated in 9, the problem of selection for this variable appears very limited provided that only 10% of the ads do not display this information. This information appears as an interesting proxy in order to control for the housing unit quality.

Table 9: Energy consumption and greenhouse gas emission

	count	mean	std	min	25%	50%	75%	max
Energy (%):A	4225940.0	3.6	18.5	0.0	0.0	0.0	0.0	100.0
Energy (%):B	4225940.0	5.8	23.4	0.0	0.0	0.0	0.0	100.0
Energy (%):C	4225940.0	13.2	33.8	0.0	0.0	0.0	0.0	100.0
Energy (%):D	4225940.0	25.6	43.6	0.0	0.0	0.0	100.0	100.0
Energy (%):E	4225940.0	17.8	38.3	0.0	0.0	0.0	0.0	100.0
Energy (%):F	4225940.0	6.2	24.2	0.0	0.0	0.0	0.0	100.0
Energy (%):G	4225940.0	2.4	15.4	0.0	0.0	0.0	0.0	100.0
Energy (%):H	4225940.0	0.0	1.8	0.0	0.0	0.0	0.0	100.0
Energy (%):I	4225940.0	0.0	1.5	0.0	0.0	0.0	0.0	100.0
Energy (%):None	4225940.0	8.9	28.5	0.0	0.0	0.0	0.0	100.0
Energy (%):V	4225940.0	4.8	21.4	0.0	0.0	0.0	0.0	100.0
GES (%):A	4225940.0	4.3	20.4	0.0	0.0	0.0	0.0	100.0
GES (%):B	4225940.0	13.6	34.3	0.0	0.0	0.0	0.0	100.0
GES (%):C	4225940.0	21.3	40.9	0.0	0.0	0.0	0.0	100.0
GES (%):D	4225940.0	13.9	34.6	0.0	0.0	0.0	0.0	100.0
GES (%):E	4225940.0	10.4	30.6	0.0	0.0	0.0	0.0	100.0
GES (%):F	4225940.0	4.8	21.5	0.0	0.0	0.0	0.0	100.0
GES (%):G	4225940.0	1.9	13.8	0.0	0.0	0.0	0.0	100.0
GES (%):H	4225940.0	0.0	2.1	0.0	0.0	0.0	0.0	100.0
GES (%):I	4225940.0	0.0	1.7	0.0	0.0	0.0	0.0	100.0
GES (%):None	4225940.0	11.2	31.5	0.0	0.0	0.0	0.0	100.0
GES (%):V	4225940.0	4.4	20.4	0.0	0.0	0.0	0.0	100.0

A.4 Bias

Table 10: Estimate of the value of the reference flat for departments and regions

	(1) ln(Price/Surface)	(2) ln(Price/Surface)	(3) ln(Price/Surface)	(4) ln(Price/Surface)
ln(surface)-ln(50)	-0.476***	-0.526***	-0.494***	-0.500***
	(0.0450)	(0.0450)	(0.0232)	(0.0511)
$(ln(surface) - ln(50))^2$	0.114***	0.0345	0.0879***	-0.00103
	(0.0190)	(0.0264)	(0.00924)	(0.0216)
1 room	-0.0313***	-0.0131	-0.0447***	0.00229
	(0.00244)	(0.0168)	(0.00343)	(0.0254)
2 rooms (ref.)	0	0	0	0
	(.)	(.)	(.)	(.)
3 rooms	0.00652	0.0118	0.0217***	0.0170
	(0.0116)	(0.0223)	(0.00513)	(0.0155)
4 rooms	0.00949	0.0249	0.0386***	0.0453**
	(0.0137)	(0.0255)	(0.00788)	(0.0190)
5 rooms	0.0205	0.0655*	0.0539***	0.0784***
	(0.0153)	(0.0368)	(0.0111)	(0.0253)
6+ rooms	-0.0286	0.0855***	0.0271	0.0893***
	(0.0372)	(0.0234)	(0.0259)	(0.0306)
Street level	-0.00972***	0.0168	-0.0118***	0.00140
	(0.00287)	(0.0210)	(0.00292)	(0.0257)
Floor 1	-0.0107	0.0226**	-0.0139***	0.0137
	(0.0117)	(0.0104)	(0.00208)	(0.0134)
Floor 2 (ref.)	0	0	0	0
	(.)	(.)	(.)	(.)
Floor 3 and 4	0.0712*	0.0565***	0.0242***	-0.0102
	(0.0364)	(0.0134)	(0.00514)	(0.0157)
$> than$ 4	0.163***	0.103***	0.0412***	0.00633
	(0.0546)	(0.0252)	(0.0110)	(0.0166)
Last Floor	0.0536***	0	0.0306***	0
	(0.00803)	(.)	(0.00305)	(.)
Extra expenditures :included	0.0894***	0.0368**	0.0847***	0.0377**
	(0.00584)	(0.0170)	(0.00325)	(0.0185)
Extra expenditures :Unknown	0.0429***	0	0.0679***	0
	(0.00930)	(.)	(0.00302)	(.)
Furnished	0.0834***	-0.0575**	0.0640***	-0.0773***
	(0.0242)	(0.0246)	(0.00663)	(0.0281)
single_unit	-0.0526*	-0.0846***	-0.0179	-0.0458*
	(0.0254)	(0.0264)	(0.0130)	(0.0256)
R2	0.693	0.637	0.789	0.698
Obs	1615070	3818	1618025	3818
Weights	Y	Y	Y	Y
Fixed Effects	REG	REG	DEP	DEP
Estimator	OLS	OLS	OLS	OLS
Data	Ads	Survey	Ads	Survey

Standard errors in parentheses

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

Authors computation online ads and the French Housing survey 2013

B The cost of density

B.1 Data Sources

variables	sources	Producer	Link
Population and Growth	Census 2015	INSEE	https://www.insee.fr/fr/statistiques/3627376
Education	Census 2015	INSEE	
Median Income	Filosofi	INSEE	https://www.insee.fr/fr/statistiques/3560121
Land Area	GeoFla	IGN	http://professionnels.ign.fr/bddvecteur
Land developed	Corine Landcover	CGEDD	https://www.statistiques.developpement-durable.gouv.fr/corine-land-cover-0
Building Height/footprint	BD Topo	IGN	http://professionnels.ign.fr/bdtopo
Equipments	Base des Equipements	INSEE	https://www.insee.fr/fr/statistiques/3568656
Geology	ESDAC	European Commission	https://esdac.jrc.ec.europa.eu/
Temperature	Global Climate Data	Hijmans et al. (2005)	http://www.worldclim.org/
Historical Census	Cassini	EHESS/INED	cassini.ehess.fr/cassini/fr/html/
Frontier	Wikipedia	Wikipedia	https://fr.wikipedia.org/wiki/
Local taxation	DGFIP/DGCL	DGFIP/DGCL	https://www.collectivites-locales.gouv.fr/fiscalite-directe

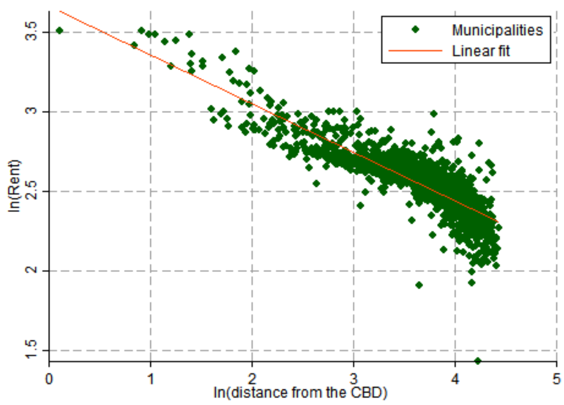
Table 11: Data source for the update and extension of control variables

B.2 Descriptive statistics

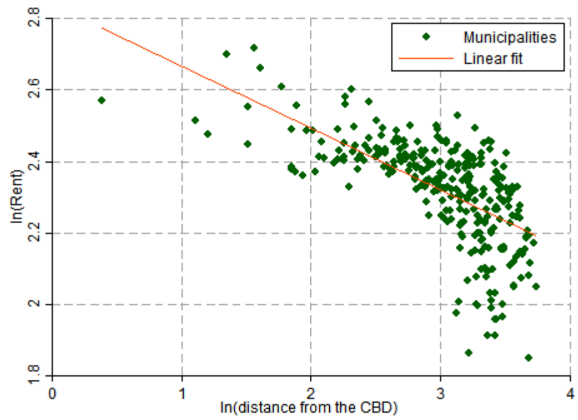
	Mean	Std.Dev.	min	p5	p50	p95	max
log(rent)	2.23	0.22	1.13	1.91	2.21	2.64	3.51
Median_income	22185.21	3228.65	11726.50	18140.67	21673.68	27841.33	46156.00
share university	23	8	0	12	22	38	73
distance	15619	13550	0.00	2976	11656	47154	83467
share artificialized	8	15	0	0	3	35	100
share footprint	1	2	0	0	0	3	61
average height	4.75	1.04	1.13	3.41	4.61	6.49	18.74
Average temperature	3.58	1.87	-8.39	0.80	3.45	6.45	9.77
hypermarche per 1000 inhabitants	0.01	0.08	0.00	0.00	0.00	0.00	4.52
supermarche per 1000 inhabitants	0.09	0.37	0.00	0.00	0.00	0.56	40.18
restaurant per 1000 inhabitants	2.17	4.99	0.00	0.00	1.11	7.52	390.24
lycee tech per 1000 inhabitants	0.00	0.07	0.00	0.00	0.00	0.00	5.81
lycee pro per 1000 inhabitants	0.01	0.16	0.00	0.00	0.00	0.00	12.20
lycee gen per 1000 inhabitants	0.01	0.07	0.00	0.00	0.00	0.00	4.98
college per 1000 inhabitants	0.05	0.19	0.00	0.00	0.00	0.32	5.81
primaire per 1000 inhabitants	0.53	0.86	0.00	0.00	0.00	2.07	15.27
primaire rpi per 1000 inhabitants	0.50	1.20	0.00	0.00	0.00	3.17	17.54
et med long per 1000 inhabitants	0.00	0.04	0.00	0.00	0.00	0.00	3.73
et med moyen per 1000 inhabitants	0.01	0.14	0.00	0.00	0.00	0.00	6.67
et med court per 1000 inhabitants	0.01	0.07	0.00	0.00	0.00	0.00	6.67
doctors per 1000 inhabitants	0.45	0.99	0.00	0.00	0.00	2.11	34.56
cardio per 1000 inhabitants	0.01	0.16	0.00	0.00	0.00	0.00	19.01
cinema per 1000 inhabitants	0.02	0.15	0.00	0.00	0.00	0.03	8.06
laboratory per 1000 inhabitants	0.01	0.07	0.00	0.00	0.00	0.09	4.52
River La Garonne	1	8	0	0	0.0	0	100
River La Seine	1	10	0	0	0	0	100
River Le Rhin	0	4	0	0	0	0	100
River Le Rhone	1	8	0	0	0	0	100
River la Loire	1	10	0	0	0	0	100
Sea	2	15	0	0	0	0	100
country Belgium	0	5	0	0	0	0	100
country Italy	0	2	0	0	0	0	100
country Luxembourg	0	3	0	0	0	0	100
country Spain	0	5	0	0	0	0	100
country Switzerland	0.00	5	0	0	0	0	100
Erodability							
0	1.82	13.39	0.00	0.00	0.00	0.00	100.00
1	1.74	13.09	0.00	0.00	0.00	0.00	100.00
2	24.64	43.09	0.00	0.00	0.00	100.00	100.00
3	24.89	43.24	0.00	0.00	0.00	100.00	100.00
4	35.02	47.70	0.00	0.00	0.00	100.00	100.00
5	11.88	32.35	0.00	0.00	0.00	100.00	100.00
Hydrogeological Class							
1	15.80	36.48	0.00	0.00	0.00	100.00	100.00
2	44.58	49.71	0.00	0.00	0.00	100.00	100.00
4	31.54	46.47	0.00	0.00	0.00	100.00	100.00
5	0.36	5.97	0.00	0.00	0.00	0.00	100.00
6	0.16	3.98	0.00	0.00	0.00	0.00	100.00
7	2.23	14.76	0.00	0.00	0.00	0.00	100.00
9	2.97	16.97	0.00	0.00	0.00	0.00	100.00
10	2.37	15.21	0.00	0.00	0.00	0.00	100.00
Dominant Material							
1	1.64	12.71	0.00	0.00	0.00	0.00	100.00
2	48.71	49.98	0.00	0.00	0.00	100.00	100.00
3	2.48	15.55	0.00	0.00	0.00	0.00	100.00
4	37.00	48.28	0.00	0.00	0.00	100.00	100.00
6	2.98	16.99	0.00	0.00	0.00	0.00	100.00
7	2.42	15.36	0.00	0.00	0.00	0.00	100.00
8	4.78	21.34	0.00	0.00	0.00	0.00	100.00

Table 12: Summary Statistics for French Municipalities

Figure 11: Rent gradients in large urban areas

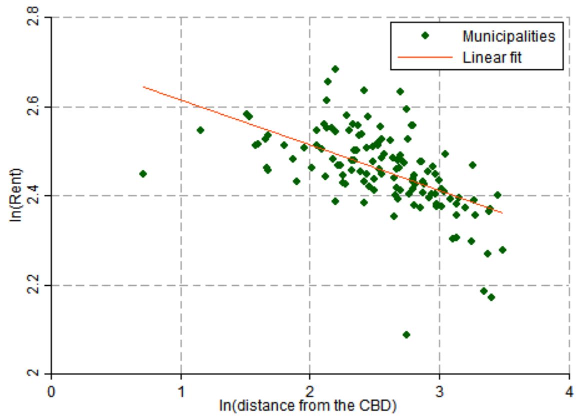


(a) Paris



(b) Lyon

(c) Lille



(d) Toulouse

	Mean	Std.Dev.	min	p5	p50	p95	max
log(rent) - no control	2.32	0.23	1.82	2.05	2.29	2.75	3.66
log(rent) - extended controls	2.25	0.20	1.86	2.00	2.23	2.62	3.49
Rent Gradient- no control	-0.06	0.06	-0.32	-0.17	-0.06	0.04	0.16
Rent Gradient - extended control	-0.04	0.10	-0.26	-0.12	-0.04	0.04	0.16
Population	140354.72	693592.73	2110.00	12303.00	37005.00	415968.00	12495784.00
Land Area	49948.86	94021.50	1297.00	3750.00	27070.50	175704.00	1456911.00
Median income	19936.45	1625.77	15111.68	17767.51	19904.68	22178.20	31424.81
Growth	0.26	0.68	-2.18	-0.88	0.26	1.37	3.09
Share university degree	23.44	5.49	11.20	15.54	22.97	32.96	41.64
mean height	5.18	1.0247	0.00	3.97	5.08	7.05	8.72
share buildings footprints	1.24	0.77	0.00	0.34	1.03	2.75	5.39
share land artificialized	12.42	7.91	1.16	3.52	10.16	28.65	54.21
Average temperature in January	3.85	2.11	-5.89	0.84	3.82	6.96	8.98
alt min	116.76	141.78	-89.67	-5.00	72.89	360.92	1008.94
alt max	525.66	638.26	21.10	65.05	273.70	1682.45	4749.67
River La Garonne	2.27	14.92	0.00	0.00	0.00	0.00	100.00
River La Seine	3.98	19.57	0.00	0.00	0.00	0.00	100.00
River Le Rhin	0.57	7.53	0.00	0.00	0.00	0.00	100.00
River Le Rhone	4.26	20.23	0.00	0.00	0.00	0.00	100.00
River la Loire	4.83	21.47	0.00	0.00	0.00	0.00	100.00
country belgique	1.99	13.98	0.00	0.00	0.00	0.00	100.00
country italy	0.85	9.21	0.00	0.00	0.00	0.00	100.00
country luxembourg	0.85	9.21	0.00	0.00	0.00	0.00	100.00
country spain	0.57	7.53	0.00	0.00	0.00	0.00	100.00
country suisse	1.42	11.85	0.00	0.00	0.00	0.00	100.00
Erodibility							
1.0000	1.14	10.61	0.00	0.00	0.00	0.00	100.00
2.0000	20.17	40.18	0.00	0.00	0.00	100.00	100.00
3.0000	26.42	44.15	0.00	0.00	0.00	100.00	100.00
4.0000	39.77	49.01	0.00	0.00	0.00	100.00	100.00
5.0000	12.50	33.12	0.00	0.00	0.00	100.00	100.00
Hydrogeological Class							
1.0000	14.77	35.53	0.00	0.00	0.00	100.00	100.00
2.0000	47.44	50.01	0.00	0.00	0.00	100.00	100.00
4.0000	30.97	46.30	0.00	0.00	0.00	100.00	100.00
7.0000	2.27	14.92	0.00	0.00	0.00	0.00	100.00
9.0000	3.69	18.89	0.00	0.00	0.00	0.00	100.00
10.0000	0.85	9.21	0.00	0.00	0.00	0.00	100.00
Dominant Material							
2.0000	52.84	49.99	0.00	0.00	100.00	100.00	100.00
3.0000	1.70	12.96	0.00	0.00	0.00	0.00	100.00
4.0000	38.07	48.62	0.00	0.00	0.00	100.00	100.00
6.0000	2.84	16.64	0.00	0.00	0.00	0.00	100.00
7.0000	0.28	5.33	0.00	0.00	0.00	0.00	100.00
8.0000	4.26	20.23	0.00	0.00	0.00	0.00	100.00

Table 13: Summary Statistics for Urban Areas

B.3 The rent price ratio

Table 14: The rent price ratio, OLS regression

	(1)	(2)	(3)
		log(rent/price)	
log(Population)	-0.110***	-0.0570***	-0.0449***
	(0.0166)	(0.0140)	(0.0130)
log(Land Area)	0.0947***	0.0635***	0.0532***
	(0.0192)	(0.0155)	(0.0144)
log(Tax rate)		0.408***	0.357***
		(0.0328)	(0.0312)
log(Price growth)			-0.213***
			(0.0308)
constant	-2.729***	-0.480**	0.206
	(0.119)	(0.204)	(0.213)
N	277	277	277
R^2	0.139	0.450	0.533

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Notes: The dependent variable is the ratio of urban area fixed effects estimated with rents and urban area fixed effects estimated with prices in Combes, Duranton, and Gobillon (2018). The superscripts a, b, and c indicate significance at 1%, 5%, and 10% respectively.

B.4 2sls Correcting for rent and predicted impact of population

Table 15: The determinants of housing rent at the city center, 2SLS regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Log price per m2, with first step and second step controls								
log(population)	0.204***	0.207***	0.204***	0.186***	0.187***	0.287***	0.201***	0.287***
	(0.0180)	(0.0177)	(0.0180)	(0.0157)	(0.0156)	(0.0472)	(0.0141)	(0.0474)
log(surface)	-0.140***	-0.142***	-0.140***	-0.114***	-0.116***	-0.218***	-0.110***	-0.218***
	(0.0162)	(0.0159)	(0.0162)	(0.0145)	(0.0144)	(0.0573)	(0.0160)	(0.0576)
N	1937	1937	1937	1937	1937	1937	1937	1937
Weak	215.9	187.2	145.4	239.0	240.4	16.3	346.4	11.0
OverId	.	0.49	0.67	0.72	0.10	.	0.05	0.72
Panel B: Log rent per m2, with first step controls and second step controls								
log(population)	0.153***	0.157***	0.156***	0.143***	0.142***	0.191***	0.137***	0.211**
	(0.0247)	(0.0246)	(0.0252)	(0.0219)	(0.0219)	(0.0674)	(0.0184)	(0.0952)
log(surface)	-0.0725***	-0.0755***	-0.0749***	-0.0565***	-0.0572***	-0.126	-0.0599***	-0.150
	(0.0159)	(0.0157)	(0.0161)	(0.0142)	(0.0142)	(0.0802)	(0.0154)	(0.114)
N	277	277	277	277	277	277	277	277
Weak	93.7	84.2	61.7	97.2	98.5	7.5	61.7	5.0
OverId	.	0.21	0.07	0.07	0.34	.	0.38	0.13
Panel C: Log rent per m2, updated population, surface and control variables								
log(population)	0.128***	0.134***	0.128***	0.118***	0.119***	0.201***	0.124***	0.191**
	(0.0264)	(0.0261)	(0.0266)	(0.0229)	(0.0230)	(0.0772)	(0.0190)	(0.0928)
log(surface)	-0.0652***	-0.0698***	-0.0654***	-0.0524***	-0.0541***	-0.143	-0.0526***	-0.131
	(0.0162)	(0.0161)	(0.0164)	(0.0143)	(0.0143)	(0.0882)	(0.0158)	(0.106)
N	277	277	277	277	277	277	277	277
Weak	76.30	67.88	50.47	73.65	73.79	6.21	35.69	4.39
OverId	.	0.12	0.24	0.25	0.30	.	0.27	0.20
Panel D: Log rent per m2, extended sample								
log(Population)	0.0908^a	0.108^a	0.0823^b	0.0769^a	0.0798^a	0.275^b	0.115^a	0.0279
	(0.0318)	(0.0297)	(0.0334)	(0.0291)	(0.0283)	(0.119)	(0.0245)	(0.178)
log(Surface)	-0.0628^a	-0.0748^a	-0.0570^b	-0.0501^a	-0.0522^a	-0.210^c	-0.0539^a	0.0296
N	352	352	352	352	352	352	352	352
Weak	51.91	56.79	37.71	56.96	52.97	1.79	30.54	2.97
OverId	.	0.03	0.25	0.27	0.24	.	0.13	0.09
Urban population in 1831	Y	Y	Y	Y	Y	N	N	N
Urban population in 1851	Y	Y	Y	N	N	N	N	N
Urban area in 1881	N	Y	N	N	N	N	N	N
Urban pop. density in 1881	N	Y	N	N	N	N	N	N
January temperature	N	N	Y	Y	N	N	N	Y
Number of hotel rooms	N	N	N	N	N	Y	Y	Y
Share of one-star hotel rooms	N	N	N	N	Y	Y	Y	Y

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Notes: The first-step controls are all the controls available. The second-step controls correspond to the controls used in columns 2, and 5 of table 8. All estimations are performed with LIML.

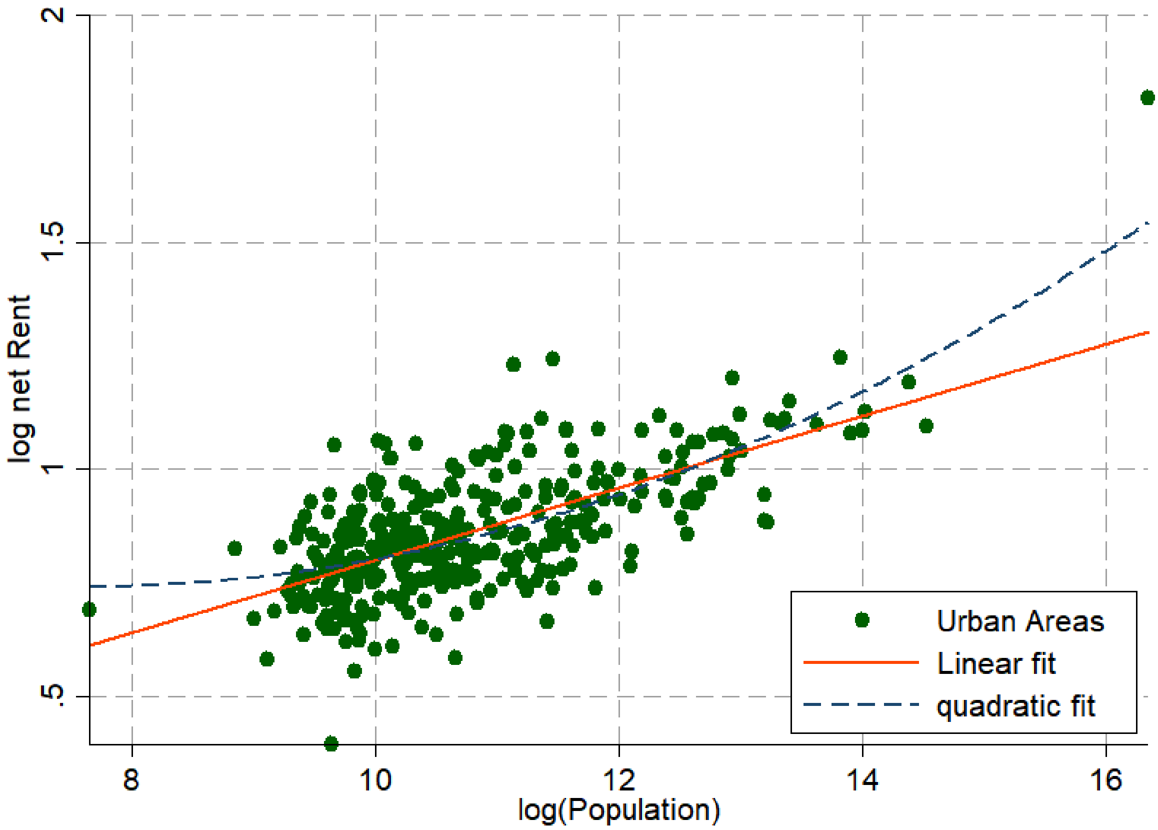


Figure 12: Residual and the predicted impact of population