Observing new leases in French Cities: evidence from the web

March 2019

1 Introduction

The literature on hedonic indices for the real estate market is relatively large. Indeed, housing price indices are increasingly widespread and available for many countries and cities. However, fine level rent indices remain scarcer. In this work, we take advantage of two large datasets on online ads to develop local rent indices for French cities. The first one comes from a partnership with one of the leading real estate web site in France, Se Loger,while the second comes from a dataset gathered thanks to webscraping as described in Chapelle and Eyméoud (2018).

This paper exposes the methodology and describes the dynamics of the rental market for several French cities. It puts particular emphasis on Paris where rent control was implemented from the second semester 2015 until the November 2017. Two major results arise from this study. First, it appears that rents in Paris were relatively stable during the rent control period and then increased dramatically when the policy was repealed. Nevertheless at this stage, it appears impossible to formally state that these patterns are caused by the policy. Second, Prices appear to increase much more rapidly than rents in the Paris, Lyon and Marseille.

2 Methodology

2.1 Hedonic models

Hedonic modeling becomes widespread to develop real estate price indices. Provided that our goal is to provide reliable rent price indices, we will follow closely the literature developed for the House price index INSEE-Notaires. We will thus estimate two complementary models following closely the exposition in Gouriéroux and Laferrère (2009) and Clarenc et al. (2014) to provide reliable rent indices.

2.2 The notaire INSEE approach

We first follow closely the methodology exposed in Clarenc et al. (2014). We first estimate correction coefficient on a reference year, in our case 2015, we then use the estimated observations to convert the observed price per square meter of online ads into the price of the reference good of the city. We first restrict our sample to the observations of the year 2015 removing observations belonging to the two uppest and lowest centiles of price per square meters within each municipality and estimate the following equation **separately for each large French municipality c**.

$$ln(p_{i,2015}) = p_0 + \sum_{k=1}^{K} \beta_{2015,k,c(i)} X_{k,i,2015} + \epsilon_{i,2015} \tag{1}$$

where p_i is the price per square meter of dwelling i observed in an online ad. X_i is the vector of observed characteristics. In our main database, which starts in 2015, we mostly have the neighborhood, the surface and the number of rooms. We use our alternate dataset which comes from webscraping for robustness checks. It starts in 2016 and also have many additional variables for each dwelling, including the exposition, the floor, the energy efficiency, the presence of a balcony etc... p_0 is interpreted as the price of a reference good (in our case, this is the good with 60 meters and two rooms, in the most observed neighborhood).

In a second step, for each period of estimation (quarters,semesters or years), we first remove the extreme observations and use convert the log price price per square meters of the remaining observations to express their price in terms of the reference good ($log(\tilde{p_{i,t}})$):

$$log(\tilde{p_{i,t}}) = log(p_{i,t}) - \sum_{k=1}^{K} \beta_{2015,k,c(i)} X_{k,i,2015} \tag{2}$$

For each city and each period, we then compute the average of $log(\tilde{p_{i,t}})$ to get an estimate of the value of the reference good

$$log(\hat{p_{0,t,c}}) = \frac{1}{J_t} \sum_{j=1}^{J_t} log(\tilde{p_{j,t}}) \tag{3}$$

The local rental price index for municipality c in period t will be the ratio of the estimated value of the reference good with respect to its value at the base period (here 2015).

$$I_t = \frac{e^{log(\hat{p_{0,t,c}})}}{e^{log(\hat{p_{0,2015,c}})}} \tag{4}$$

This method has the strong added value to be easy to update as corrections coefficients are estimated with data of the period before the index. For the INSEE-Notaires indices, the hedonic model is updated every two years the index of the year t and t+1 are computed with coefficients computed with the data of the year t-2 and t-1. On the long run, we aim to follow the same methodology.

2.3 The dummy approach

The INSEE-Notaires approach relies on the assumption that the marginal contribution of hedonic characteristics are similar between the period of estimation of the hedonic model (t-1 and t-2) and the period of the indices (t and t+1). Following Balcone and Laferrère (2018) and assuming that $\beta_{t,k,c(i)}$ is constant through all the periods, an alternate model can be used : the dummy approach. This alternate specification is usually popular among scholars but of relatively little use to publish regularly updated indices as for each iteration, correction coefficients are reestimated affecting the estimated change of previous periods. Nevertheless, as a robustness check, we propose to confront our index with this alternate approach. In this specification, the following model is estimated on all the dataset excluding extreme deciles computed within periods:

$$log(p_i) = log(p_0) + \sum_{k=1}^{K} \beta_{2015,k,c(i)} X_{k,i,2015} + \sum_{t=0}^{T} \delta_{t,c} \times 1_{t(i)=t} + \epsilon_i \tag{5}$$

where p_i is the price per square meter in the online ads. $1_{t(i)=t}$ is a dummy indicating whether the ad was published in period t while T is the total number of periods. In such a case the index is easily recovered as it is simply:

$$I_{t,c} = e^{\delta_{t,c}} \tag{6}$$

3 Data

3.1 Se Loger Dataset

	count	mean	std	min	25%	50%	75%	max
Rent	3310813.0	793.0	612.4	30.0	499.0	650.0	856.0	14900.0
Rent per square meter	3310813.0	15.7	8.4	2.0	10.0	13.1	18.9	100.0
surface	3310813.0	56.5	32.4	5.4	35.0	50.0	70.0	3800.0
Rooms (%):01	3310813.0	21.2	40.9	0.0	0.0	0.0	0.0	100.0
Rooms (%):02	3310813.0	34.1	47.4	0.0	0.0	0.0	100.0	100.0
Rooms (%):03	3310813.0	26.3	44.0	0.0	0.0	0.0	100.0	100.0
Rooms (%):04	3310813.0	11.6	32.0	0.0	0.0	0.0	0.0	100.0
Rooms (%):05	3310813.0	4.5	20.8	0.0	0.0	0.0	0.0	100.0
Rooms (%):6+	3310813.0	2.3	14.8	0.0	0.0	0.0	0.0	100.0

Table 1: Descriptive statistics for the Se Loger dataset

We use the dataset stored by Se Loger. This dataset is the universe of ads that were posted on the website since 2015. Unfortunately, the number of variables available remains limited as the text was deleted. We only know the price, the surface and the number of rooms and the location of the good. Table 1 describes the main variable available. There is an over-representation of some large cities as Paris or Lyon in the datasets where the website Se Loger is extremely popular. The yearly number of ads appears to be relatively constant.

3.2 Web scraping

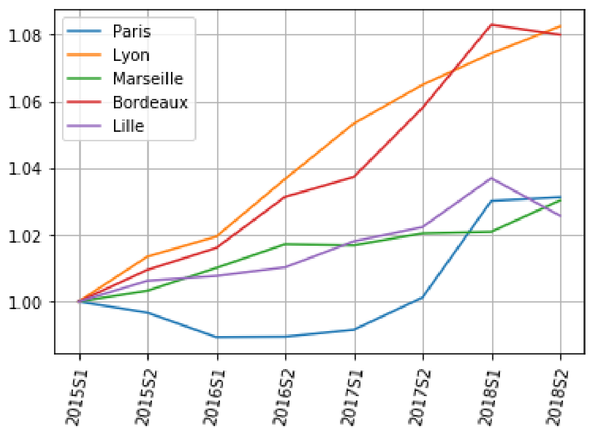
Alternatively we also use the dataset gathered and extensively presented in Chapelle and Eyméoud (2018) that has been continuously updated until nowadays. This dataset is based on the webscraping of the major French Real estate website starting at the end of 2015. There is thus one year less than what is in the Database of Se Loger. However, this method allows to gather all the information on each good and thus to increase dramatically the number of controls. We can get the amount of extra expenditures, the status (furnished or not), the presence of a balcony or a garden, the floor using the information directly coded in the web page or thanks to regular expressions in the description.

4 Results

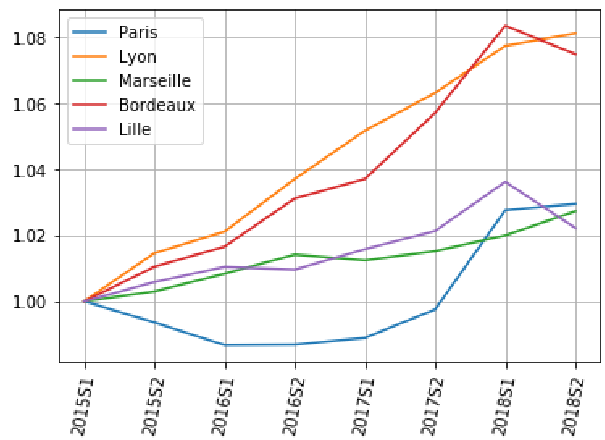
4.1 Rent trends for five major Municipalities

We apply our method for the large municipalities in Figure 1. One can observe that both methods yield extremely close results while the gross average tend to overestimate rent increase. On striking fact is the behaviour of Paris which went through a slight decline and stagnation of the price of new lease followed by a sharp increase in Rent price from the second semester 2017 which correspond to the end of the rent control. On can also compare the evolution of its rent with its close suburbs in panel c); while Paris went through a slight decline, its suburbs remained stable until the second semester of 2017 where rent increased for all the four departments. This patterns might suggest that rent control contributed to moderate rent price dynamics in Paris and its close suburbs but that decontrol led to a readjustment leading rents to similar cumulated growth as a city like Marseille. It is worth noting that one does not observe a similar break in Lille where rent control was only applied for a very short period of time. Infering any causal implication of rent control on the observed patterns in Paris remain hard one should find a credible counterfactual to try to assess how would have prices increased in Paris.

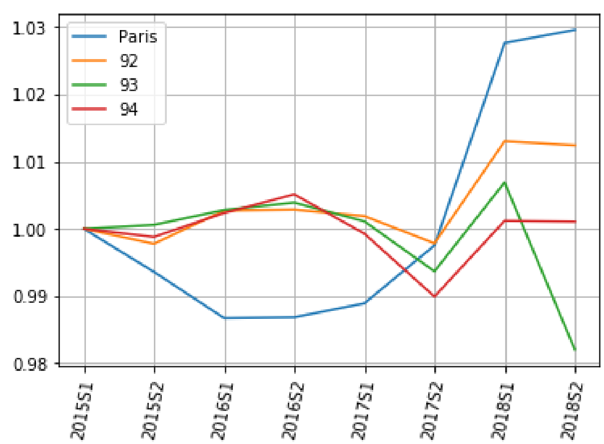
Figure 1: Rent indices with alternate methods for five French municipalities



(a) Rent indices with the INSEE-Notaires method



(b) Rent indices with dummy method



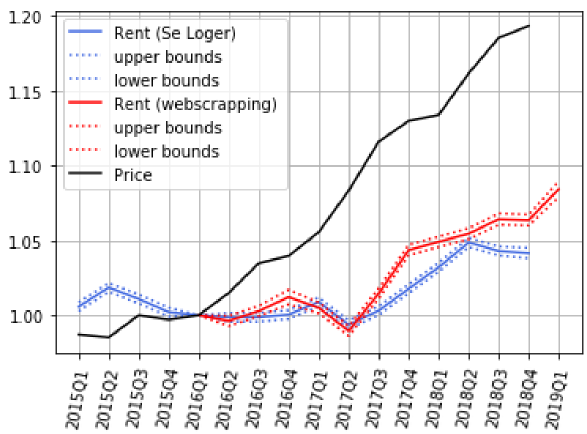
(c) Rent indices for Paris and its close suburbs

Controls: Municipal (and arrondissements) fixed effects, log of the surface, number of rooms, and type of unit (single Vs flat)
Source: Author's computation from the Se Loger dataset

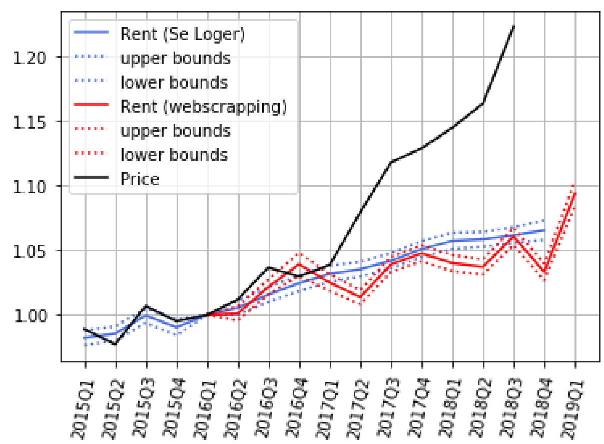
4.2 Comparison with price and alternate data sources

We confront the dataset from Seloger with the dataset from webscraping and the evolution of housing price indices provided by the INSEE for Paris or the agglomeration of Marseille and Lyon. One can observe that trends between the two rent series are extremely similar. Some divergences might arise as the webscraping also cover another website with a large amount of ads posted directly by the landlords which appear to be more volatile. It is also worth noting that the dramatic increase in rent price from the last semester of 2017 in Paris is observed in both datasets. Moreover, one can observe a steep surge in housing price much more pronounced that what is observed in the rental market.

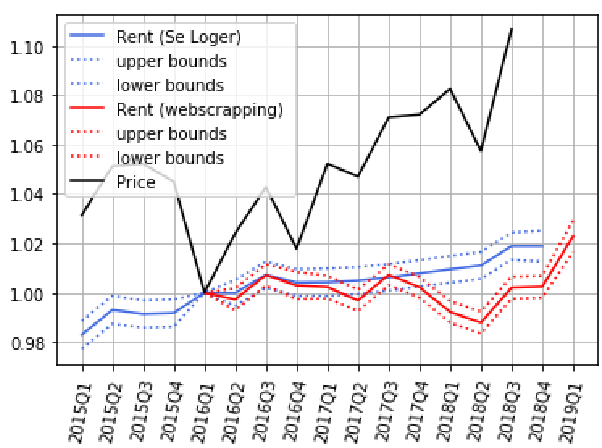
Figure 2: Alternate data sources and comparison with prices



(a) Paris



(b) Lyon



(c) Marseille

Controls with Seloger dataset: Municipal (and arrondissements) fixed effects, log of the surface, number of rooms, and type of unit (single Vs flat)
Controls with the webscrapping dataset: Type of units, surface, number of rooms, type of contract, presence of a kitchen, floor. Inclusion of extra expenditures or not in the price.
Source: Author's computation from the Se Loger dataset and webscraping database
Price indices are the INSEE-Notaire gross indices for flat for Paris, Lyon Urban area and Marseille Urban Area

5 Conclusion

We present a new dataset that could potentially increase our knowledge of the dynamics of local rental markets. When comparing this dataset with the one gathered in Chapelle and Eyméoud (2018), dynamics appear very similar even if some discrepancies might arise from the fact that Chapelle and Eyméoud (2018) also cover ads directly posted by landlords without intermediary. The limited number of controls in the Seloger Dataset does not change dramatically the observed trend when the number of observations is large as for cities like Paris, Lyon or Marseille. In these cities, rents appear relatively dynamic, in particular in Bordeaux and Lyon where rents increased by almost 8% between 2015 and the end of 2018. Nevertheless, it is worth noting that thee growth rates still appear much less dynamic than prices in these large urban areas.

References

Balcone, Thomas, and Anne Laferrère. 2018. "New or old, why would housing price indices differ? An analysis for France." *Economie et Statistique* 500 (1): 69–95.

Chapelle, Guillaume, and Jean-Benoît Eyméoud. 2018. "Can big data increase our knowledge of local rental market?" *Sciences Po mimeo.*

Clarenc, Philippe, Jean-François Côte, Alain David, Jacques Friggit, Philippe Gallot, Stéphane Gregoir, Anne Laferrère, Adélia Nobre, Catherine Rougerie, and Nelly Tauzin. 2014. "Les indices Notaires-INSEE de prix des logements anciens."

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