CENTER FOR RESEARCH
IN ECONOMICS AND STATISTICS

Série des Documents de Travail

n° 2017-13
From Latin Quarter to Montmartre
Investigating Parisian Real-Estate Prices

M.POULHES[1]

Les documents de travail ne reflètent pas la position du CREST et n'engagent que leurs auteurs.
Working papers do not reflect the position of CREST but only the views of the authors.

[1] INSEE; CREST. E-mail : Mathilde.Poulhes@ensae.fr

From Latin Quarter to Montmartre Investigating Parisian Real-Estate Prices*

Mathilde Poulhes
INSEE-CREST

November 6, 2017

Abstract

This paper estimates buyers' preferences for dwelling attributes and neighbourhood characteristics. The collected data allows for the simultaneous consideration of a wide range of intrinsic characteristics, such as surface, floor, etc., and neighbourhood characteristics, including noise, crime, school quality, distance to jobs, etc. The marginal willingness to pay is identified from transaction data under the assumptions of the hedonic model described by Rosen (1974). We use very local fixed effects combined when possible with administrative boundaries as geographical discontinuities to isolate the effect of each amenity. Estimation is achieved by using flexible semi-parametric methods. Characteristics explain more than 90% of the variance of dwelling prices, showing a significant positive marginal willingness to pay for job accessibility and school quality, and a negative marginal willingness to pay for a higher crime rate in the area. By contrast, noise level or public transport accessibility have less influence on housing prices. These results are robust to the inclusion of census tract fixed-effects, which also drastically reduces the spatial correlation of the residual prices.

*I am grateful to Nathaniel Baum-Snow, Odran Bonnet, Pierre-Henri Bono, Paul Cheshire, Denis Fougère, Alfred Galichon, Hans Koster and Maxime Tô. I benefited from the comments of the seminar participants of the 6^{th} European Meeting of the Urban Economics Association and of the *2015 Commission des comptes du logement*. The views expressed in this article do not necessarily reflect those of INSEE.

I Introduction

The average price per square meter of flats sold in Paris in 2008 varied by a factor of two. Moreover, even within each Paris arrondissement[1], prices experienced large variations. Part of this local variation can be associated to differences in dwelling characteristics, but not all. The remaining price differences measure the desirability of the location.

In this paper, we take advantage of a very detailed transaction data that gathers a wide range of intrinsic dwelling characteristics. It also gives the precise location of the property, so we are able to estimate the housing-price premia attached to the following local factors: noise, crime, school quality, distance to jobs and distance to public transportation. These factors are often cited among the most important recent urban problems, and focus the attention of many urban researchers. They are also clearly identified as crucial factors in the formation of real-estate prices by the urban economic literature. Gibbons & Machin (2008) review the main results concerning the role of school, crime and transport in generating housing price variations. Ahlfeldt & Wendland (2016) focus on the importance of job accessibility, and Nelson (2004) looks at the role of noise focusing airports only. Other local dimensions like air pollution could have been taken into account, but we limit our analysis to these dimensions for which we have relevant data.

The valuations of these amenities are also of great interest for policy makers. A better understanding of residents' preferences should improve public policy-making. Because we consider simultaneously the different factors, we can directly compare the willingness to pay for the different dimensions and establish a hierarchy between the valuations of the pre-cited public goods. Once adjusted for variations in dwelling characteristics, housing prices variability reveals relevant information about the marginal willingness to pay for accessibility, school quality and other neighbourhood attributes.

Our theoretical framework is the one of hedonic prices. Each dwelling is considered as a good differentiated by its characteristics (surface, quality of neighbourhood, distance to jobs, distance to public transportation, etc.). The hedonic price function relates a price to a bundle of characteristics, and determines the implicit hedonic price of each attribute of the dwelling.

Court (1939) is known to be the first to use this framework in order to compute quality corrected price indexes. At the end of 1960s, Lancaster (1966) and Rosen (1974) established a theoretical framework. They were the first to interpret the hedonic price function as a result of the equilibrium between supply and demand. At this time, hedonic techniques

[1]Paris is divided into 20 municipalities called "arrondissement". On average one arrondissement approximately covers 5 km^2

appeared as a powerful tool to identify structural parameters of supply and demand on a differentiated good market, and the method was widely used as such.

During the 1990s, the hedonic approach experienced a decline. Brown & Rosen (1982) and Bartik (1987) found identification problems in the method of Rosen (1974). In particular, the simultaneity of the supply and demand equations implies a systematic bias in the two-step estimation proposed by Rosen. Moreover, the new interest for the imperfect competition markets thanks to the tools developed by Berry et al. (1995) drew attention away from the hedonic approach, which is less suitable for the study of these types of markets. In the 2000s, new identification results (Ekeland et al. 2004 and Heckman et al. 2010) and new estimation methods renewed the interest in the hedonic method.

But, as mentioned by Gibbons & Machin (2008), the recent focus of most applied empirical work on micro housing models has shifted away from attempting to estimate demand function parameters and is now on proper estimation of the equilibrium implicit prices. Natural experiments or regression discontinuity designs are used to address the endogeneity problem. For instance, Chay & Greenstone (2005) use a change of legislation on the authorized concentrations of total suspended particles to estimate the capitalization of air pollution into housing values while Boes & Nuesch (2011) exploit an unexpected change in flight regulations to assess aircraft noise effects. For the influence of public transportation access, Gibbons & Machin (2005) implement a quasi experimental approach by using the construction of new metro stations in London in the late 1990s. Lastly, many papers use boundary discontinuity regression to estimate school effects. Yinger & Nguyen-Hoang (2016) and Machin (2011) summarize the recent results on this particular topic: the most reliable studies suggest small capitalization effects (below 4%) in many developed countries like Canada, the United States, the United Kingdom, New Zealand, France or Norway.

As far as we know, only few papers estimate implicit prices of neighbourhood characteristics in France. Cavailhès (2005) use the French housing survey of 1996 to estimate the effect of different dwelling and neighbourhood characteristics on rents but the measures of neighbourhood quality are rather rough. Exploiting the data of the Parisian notaries, Trannoy et al. (2007) use the hedonic method to assess the effect of a neighboring renovation in Paris, Goujard (2011) investigates the impact of the presence of social housing on the sale prices of private housing, and Fack & Grenet (2010) derive parents' school quality valuation by developing a matching framework to compare sales across school attendance boundaries.

In this paper, in line with this recent literature, we do not intend to recover the demand parameters but we estimate the marginal willingness to pay of the highest bidder for each price determinant. We contribute to that literature by using a discontinuity design with administrative boundaries when it is possible and by introducing very local measurements of neighbourhood attributes (noise, crime, school quality, etc.) in our estimations in order to directly compare the valuations for these different characteristics. Finally, we show that the treatment of spatial correlation, recommended by Can (1990), is useless when we include enough data to describe the housing environment.

In the first section, we define the theoretical framework that allows us to interpret the hedonic prices, and we acknowledge the necessary assumptions. Section 2 introduces our data. Finally, we present our results in section 3. The last section concludes.

II The Hedonic Price Function

II.1 Market Structure and Competition Type

In a differentiated good market, the equilibrium transaction prices depend on the competition type of the market, the preferences of buyers and on the costs of sellers. Hedonic prices, that link prices to characteristics, are not directly informative about demand or supply. Under the assumption of imperfect competition where prices are determined by a Nash equilibrium, Pakes (2003) shows that the hedonic regression parameter is equal to the marginal cost of production plus a mark-up term related to the market power of the seller that also depends on the preferences of the buyers. In other terms, the coefficients of the hedonic regression depend on the demand elasticity.

In such a situation, the hedonic prices of the different characteristics are not interpretable *per se*. Pakes (2003) develops the example of the automobile sector in which the market power of the firms is usually high. In this case, there is no reason to think that the hedonic prices of characteristics should be stable over time, nor to think that the sign of the hedonic price should be consistent with the valuation (positive or negative) of this characteristic by the buyers. This could explain the counter intuitive results obtained by several studies on automobile or computer markets (Hulten 2003).

Nevertheless, it seems reasonable to assume that the sellers have no market power in the case of the residential real-estate market. The majority of the sellers are households (88% in our database), and their behaviour is not likely to have an impact on equilibrium prices. Assuming identical atomicity for buyers, we can relate the hedonic regression coefficients with the marginal willingness to pay at the equilibrium. We develop a simple model in the next paragraph to prove this assertion (see Nesheim 2006 for details).

II.2 Identification of Preferences from Equilibrium Hedonic Prices

We consider a real-estate market where supply and demand have no market power. Supply is fixed. Each dwelling proposed on the market is defined by a vector of characteristics $z \in Z_m \subseteq \Re^{n_z}$. The hedonic price function associates a price $p(z)$ to each vector z. Buyers are characterized by a vector $x \in X \subseteq \Re^{n_x}$ of individual characteristics, which contains their income and other attributes that could influence preferences (level of education, age, number of children, etc.).

Conditional on the price function $p(z)$, the buyer chooses the dwelling that maximises her utility:

$$\max_{z \in Z_m} U(x, z, p(z)) \qquad [1]$$

The budget constraint is implicitly included in $U(x, z, p(z))$ since the type x of the buyer includes income x_I. Her non real-estate consumption and savings are then $x_I - p(z)$.

The solution of this maximisation problem gives us the demand hedonic function $z = d(x)$, which matches individual characteristics to dwelling characteristics.

Under the assumption that Z_m is a compact convex subset of $\Re^{n_z}$, and that U and p are differentiable, the problem has an interior solution. The first order condition with respect to a characteristic z_j is:

$$\frac{\partial p(z)}{\partial z_j} = -\left(\frac{\partial U(x, z, p(z))}{\partial z_j}\right) / \left(\frac{\partial U(x, z, p(z))}{\partial p}\right) \qquad [2]$$

At the equilibrium, the implicit price of z_j is equal to the marginal rate of substitution of the consumer x who chooses quantity z_j (this consumed quantity at the equilibrium z_j is affected by x through the demand function $z = d(x)$). To directly interpret the hedonic price in terms of marginal willingness to pay, we have to assume the utility to be quasilinear. The maximization program of the buyer is then:

$$\max_{z \in Z_m} u(x, z) - p(z), \qquad [3]$$

and the first order condition simplifies to:

$$\frac{\partial p(z)}{\partial z_j} = \frac{\partial u(x, z)}{\partial z_j}. \qquad [4]$$

The marginal price of a characteristic is then equal to the marginal utility of the buyer. From equation [4], we can distinguish several difficulty when estimating the hedonic price function. Indeed, in equation [4] the price depends on the buyer's type x (income, level of education, number of children, etc.) which influences the choice of the real-estate property z. So the slope of the hedonic price curve only provide local information on marginal willingness to pay at the equilibrium point z, and this information is potentially heterogeneous by individual type x.

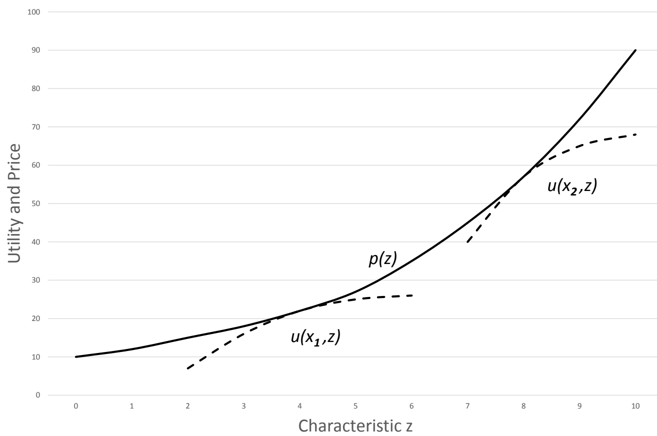


Figure 1: Hedonic Price Function and Buyers Utilities

Figure 1 illustrates this point. In this particular case, we consider two different buyers 1 and 2, and dwellings characterized by a unique characteristic (for instance the area). Buyers choose their optimal level of characteristics z according to their preferences and to their budget constraint: we note these levels z_1 and z_2. In z_1, the marginal price of the surface is equal to the marginal willingness to pay of the buyer of characteristics x_1. Similarly, in z_2, the marginal price is equal to the marginal willingness to pay of the buyer of characteristics x_2. And we see from the plot that the marginal price in z_2 is higher than the marginal willingness to pay of the individual of characteristics x_1.

Therefore, the analysis of the price curve allows us to estimate the marginal willingness to pay of the buyers at the equilibrium for the different characteristics of the real-estate property. It should be highlighted that the analysis of hedonic prices is local, and gives no indication of the reaction of the buyers to a substantial change of the price curve. Such counterfactual exercises require the estimation of deeper preference parameters of the buyers. To do this, we need to observe the characteristics of the buyers and to make some assumptions described in Ekeland et al. (2004) to ensure the identification of the model. This is left for future research.

To sum up, our estimation of the hedonic price function gives only an indication of the variation of utility of the buyers when a characteristic experiences a marginal change. More precisely, we estimate for each value of the characteristic the marginal willingness to pay of the highest bidder.

Given the dwelling characteristics that we observe in our data, it is useful to develop the interpretation frame for discrete characteristics (for instance the floor, etc.) in hedonic price functions. When assuming quasilinear utility function, the equilibrium is now defined by a set of inequalities. If the buyer chooses between J alternatives $(z_1, z_2, \ldots, z_J)$ with corresponding prices $p_j = p(z_j)$, then the preferred alternative z_i is defined by:

$$u(x, z_i) - p_i \geq u(x, z_k) - p_k \; ; \forall k \in \{1, \ldots, J\} \qquad [5]$$

The difference between the prices of the two alternatives i and j should be interpreted carefully. If the buyer is indifferent between i and j, the inequality [5] is binding for $k = j$: the difference in prices exactly compensates the difference in utility provided by the two alternatives. If the inequality is not binding, the inequality [5] is strict, and the difference in prices only provides a lower bond of the difference in preferences. As in the continuous case, nothing about the buyer's reaction to a substantial change in prices of discrete characteristics can be said.

II.3 Estimation

The previous results show that under the assumptions of atomicity and quasilinearity of the utility function of the buyers, the hedonic prices of the continuous characteristics identify the marginal willingness to pay of the observed buyers at the equilibrium, and the hedonic prices of the discrete characteristics identify a lower bound of this marginal willingness to pay. We now discuss the issues relative to the estimation of the price function $p = f(z_1, \ldots, z_K)$ relating the price of the dwelling to its characteristics.

The first issue is to determine the functional form of the price function. The equilibrium hedonic model described above does not impose any constraint concerning the functional form of the hedonic price function $p(z)$. Ideally, this hedonic price function should be estimated non-parametrically. But in practice, the number of characteristics is high relative to the sample size, and non-parametric estimation becomes difficult. Different parametric forms for the dependence between the price and the characteristics coexist in the literature: Box-Cox, semi-log, log-log. In our case, we follow here the recommendations of Diewert (2003), and we use a logarithmic transformation of the price, and suppose additivity of the unobserved heterogeneity term ε:

$$\log(p) = f(z_1, \ldots, z_K) + \varepsilon \qquad [6]$$

Regarding the functional form of f, we try to be as flexible as possible. In particular, we apply semi-parametric methods *à la* Robinson (1988) for the floor area, which is achieved by using spline regressions (for more details on the sieves method and on its properties, see Chen 2007).

Second, the noisy observation of the dwelling characteristics may be a source of bias. Indeed important unobserved price determinant may cause omitted variable bias (see Nesheim 2006 for details). Several examples from the literature illustrate this problem. Chay & Greenstone (2005) review this problem for the valuation of air quality: some papers find a positive price for fine particles concentration because they omit to control for the precise location: air pollution is often located in urban areas where prices are higher because of other amenities (see Smith & Huang 1995).

Instrumental variable is a classical solution for such a problem. Chay & Greenstone (2005) use the changes in air pollution regulation as an exogenous source of variation for the concentration of suspended particles in US counties. Bajari & Benkard (2005) suggest another solution which consists of assuming that only one characteristic is unobserved. Under this assumption, they propose a method to estimate this unobservable variable under monotonicity assumption. But the interpretation of this variable is proving difficult for the real-estate market.

Finally, the most frequent way to eliminate unobserved local effects is to work with differences between close dwellings. This is a particularly attractive identification strategy when there exist well defined boundaries in amenity quality.

In our estimates, we rely on that last identification strategy. In addition to the sets of controls that we observe (greater proximity to employment areas, to public services, etc.), we include census tract fixed effect, that accounts for unobserved determinants of prices at a very local level (a census tract in France is about .6 km^2, but this area reduces to about .3 km^2 for urban areas, and .1 km^2 in the Paris municipality.). When this is possible, as in the case of school quality, we will exploit geographical discontinuities.

III Data

III.1 Geographical Scale

In this study, we are interested in Paris city exclusively which is a relatively homogeneous urban area. Moreover, each transaction is geocoded, so we can control for the characteristics of the neighbourhood, or introduce census tract (IRIS[2] see Figure 2) fixed effects. To take into account for spatial correlation, we cluster standard errors at the IRIS level. There are 992 IRIS in the Paris centre. One IRIS corresponds to an average surface of 0.1 km^2. When we are interested in the hedonic price of a variable which is measured at the IRIS level, we introduce fixed effects at the level of the Grand-Quartier (we count 80 *grands-quartiers* in Paris, see Figure 2).

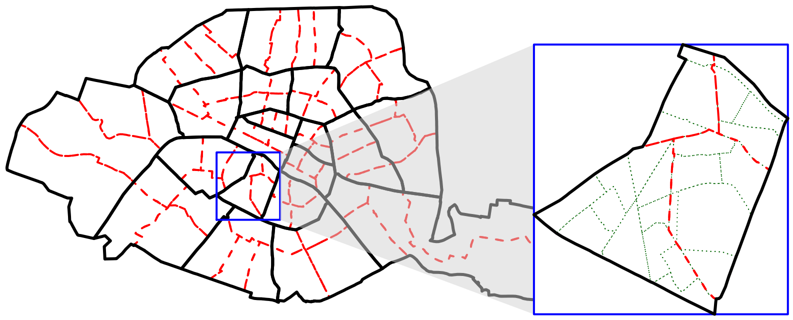


Figure 2: Parisian Administrative Districts

Notes : Continuous boundaries correspond to the limits of the arrondissements, dotted lines correspond to the limits of the *grands quartiers* and dashed lines correspond to the limits of the IRIS. In the blue box, we focus on the $VI^{\text{ème}}$ arrondissement.

III.2 Housing Unit Intrinsic Characteristics

Our main data source is the *Bien* database of the Parisian notaries[3]. The *Bien* database gathers all the real-estate transactions in the Paris region. A comparison with administrative data on transfer taxes paid shows that only 87% of transactions were reported in 2005. We restrict our analysis to flats sold between 2008-2010, since houses are scarce in Paris, and should be considered as very specific properties. Moreover, even if transaction data are available since 2004, most of the control variables we use are only available for these more recent years[4].

For each transaction, the dataset includes information on price and dwelling characteristics: surface, number of rooms, date of construction. The precise location (geographical

[2]IRIS is an acronym of "aggregated units for statistical information", it refers to a neighbourhood of about 2000 individuals and is the smallest census tract unit in the French Census.

[3]The access to these data is governed by an agreement between GENES and Paris Notaires. Source: Notaires Paris-Region, Base Bien, for the 2008-2010 period.

[4]Using 2008 crime data to estimate hedonic prices in a regression with 2006 transactions is likely to create a reflexion bias. People who bought a flat in 2006 may have influenced 2008 crime rates but we want to assess the impact of crime on real-estate prices and not the reverse.

coordinates) of each transaction is also provided. We also observe the rental status of the dwelling (type of ownership) at the time of sale: the dwelling can either be "free" of any rental contract (91.4% of the sample), "occupied by the buyer" meaning that the buyer is the former tenant (2.2%), "occupied by a tenant" referring to a situation where a tenant who is not the buyer resides in the dwelling (6.0%). A residual category ("Partial occupation") refers to the case where only part of the property is rent (0.4%). We also control for building built less than five years before the transaction since these dwelling are also subject to VAT (1.1% of the sample). Finally we control for seasonality with month dummies, and for the general evolution of housing prices with year dummies. See Ngai & Tenreyro (2014) for the existence of seasonal variations in housing prices.

Figure 3: Number of Transactions between 2008 and 2010 by IRIS Normalized by the Size of the Private Housing Stock in the IRIS (except social housing units).

Source: Bien database and Census, Insee

Figure 3 represents the distribution of transactions by IRIS from 2008 to 2010. Darker areas indicate higher number of transactions relative to the private housing stock in the IRIS. White areas show the natural outlines of Paris: the river *Seine*, and several parks: the *Champ de Mars*, the *Buttes-Chaumont*, etc. The heterogeneity between IRIS is wide, ranging from less than 2% to more than 20% of the housing stock. The real-estate market of the right bank (north part) seems more dynamic than the left bank market. Since we cannot identify a specific flat in the database, it is not possible to detect if a flat is sold for the second or third time. Therefore we ignore the exact proportion of flats that change ownership in three years.

Table I presents some descriptive statistics of the sample used in this paper. The average price per square meter decreases slightly between 2008 and 2009, and increases in 2010,

Table I: DESCRIPTIVE STATISTICS OF THE SAMPLE

	Price per square meter	Surface	Number of rooms	Number of transactions
2008	6635.42 (1926.81)	50.58 (35.64)	2.39 (1.28)	18,758
2009	6350.38 (1821.96)	51.76 (35.58)	2.43 (1.28)	16,107
2010	7056.48 (1988.67)	53.91 (36.95)	2.48 (1.31)	19,530
Total	6702.19 (1940.69)	52.12 (36.12)	2.43 (1.29)	54,395

Source : Base Bien

exceeding its 2008 level in accordance with the official Insee-Notaires price index[5]. The sales volumes experience the same trend. The average surface of a flat purchased during this period is around 52 square meters and the average number of rooms is approximately 2.5.

III.3 Local Characteristics

From the hedonic model we make the assumption that the buyer observes the distributions of flats and neighbourhood characteristics to make her decision. Thus we have to combine the characteristics of real-estate transactions with data on economic context, and neighbourhood quality. In what follows, we describe the different characteristics we use in our estimation.

Neighbourhood social characteristics: We call neighbourhood social characteristics all the variables defined at the IRIS level that describe the socio-economic context. These variables are constructed at the IRIS level, so discarded when fixed effects are introduced.

We do not directly focus on the effect of these variables, but use them as controls for neighborhood “social quality” that we want to disentangle from other location features we are interested in (noise, crime, etc.). These covariates are mainly built from the 2006 Census data (level of education, type of construction, age distribution, etc.). We also take advantage of fiscal data (*Dispositif Revenus Localisés*) to compute the median of the taxable income distribution in each IRIS. The level of public amenities in the neighbourhood of the transaction is obtained from the *Base Permanente des Équipements* database from Insee. It provides the number of shops, and restaurant by IRIS. All the control variables built from these three sources are described in Table XI of appendix B. In the result tables, we indicate the use of these control variables by "neighbourhood characteristics".

[5]Insee (the French National Institute for Statistics and Economic Studies) and the National Union of Notaries compute a price index for dwellings located inside the Paris region thanks to a hedonic method applied to the Bien database.

Distance to public transport: We compute the minimum distance between each transaction and a public transport network entry by using the geographical coordinates of all the metro and RER (commuter trains) stations provided by national rail company (SNCF) and the major transport operators in the Paris region (RATP).

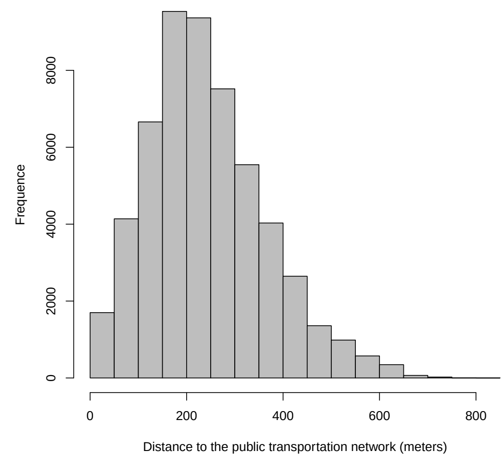


Figure 4: Minimal Distance to the Public Transportation network

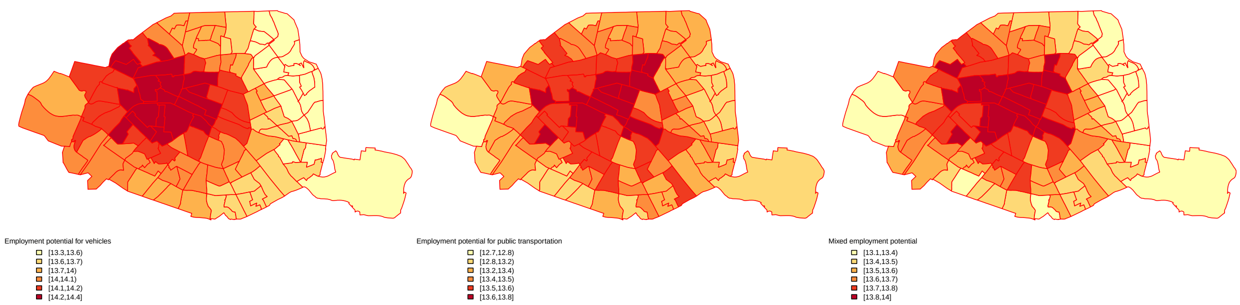
Source: Bien database and RATP and SNCF databases.
Notes: We compute the minimal euclidean distance between each transaction and the metro stations.

Figure 4 shows the distribution of the minimal distance between transactions and transport public network. Less than 1% of the transactions are located more than 600 meters from public transport and around 50% are located less than 230 meters.

Job accessibility: In line with the recent empirical literature (see Ahlfeldt & Wendland 2016 for details), we compute for each zone an employment potential to capture the labour market accessibility. Equation [7] details the computation of the employment potential. The employment potential (EP) of a given location (i) is the sum of employment (E_j) across all potential commuting destinations (j), weighted by the transport costs (t_{ij}). The weights are adjusted by a spatial decay (δ) common to all locations that determines the spatial discount according to the standard exponential cost function. Spatial decay may be estimated but the existing studies tend to find similar values. Here we use a decay parameter of 0.05, which is the parameter estimated by Ahlfeldt (2013) for London. We restrict the accessible jobs for Parisian residents to jobs located in the Paris region.

$$EP_i = \sum_j E_j \exp(-\delta \times t_{ij}) \qquad [7]$$

To compute this index, we combine data from Insee and data from DRIEA (the regional territory planning agency called *Direction Régionale et Interdépartementale de l'Équipement et de l'Aménagement d'Ile-de-France*). On the one hand, the CLAP data (*Connaissance Locale de l'Appareil Productif*) made by Insee gives us the number of jobs for every IRIS in the Paris region. This very detailed photography of the geographical distribution of jobs is only available for the year 2009. On the other hand, the DRIEA database provides the commuting times between some predefined zones used in the MODUS model, which is the regional transportation planning system of the Paris region. Each zone of the MODUS model corresponds on average to nine IRIS for Paris. Because the zones are constructed by aggregation of IRIS, our job accessibility measure has no within IRIS variability. Nonetheless, the remaining intra-IRIS variability should be low since it only concerns jobs located in the IRIS, those are few compared to the total number of jobs in the Paris region.



(a) With personal vehicle

(b) With public transportation

(c) Weighted potential, 75% public transportation, 25% personal vehicle

Figure 5: Employment Potential

Source: Insee, DRIEA and authors' calculations.

Two types of commuting times are available considering whether people commute using public transport, or personal vehicles. Both are highly correlated. To construct a unique measure of job accessibility, we use a weighted potential which combines these two measures. We choose the weights following the results of Bidoux et al. (2017) on commuting modes of the Parisian population: excluding non-commuters and walking commuters, 75% of the Parisian workers commute by public transport and 25% by road (bike, motorcycle or car).

Figure 5 represents the logarithm of the three employment potentials: by car, by public transportation and the weighted one. The west of Paris has better road access to the jobs located in the Paris region than the east (figure 5a). This is due to both a west distorted jobs distribution and a better road network in western suburbs.

The center of Paris is better connected to jobs in public transportation compared to the rest of the city (figure 5b). This is partly due to the proximity to commuting train stations at the very center of Paris. Nonetheless, the employment potentials with public transportation are weaker than the employment potentials with personal vehicle. Finally, the weighted potential (figure 5c) is by definition a mix of the previous results, showing high values for the center and the west of Paris.

Crime data: We also match the transaction location to crime activity in the neighbourhood. Crime is measured from the 2008 Police data[6]. This unique data records any crime or misdemeanours (*crimes et délits* in the French legislation) for which someone filed a complaint with the police. For each infraction, the precise address where the incident took place is recorded as well as the nature of the infringement committed. We report in the appendix A the spatial densities of the different types of infractions.

For each transaction, the number of crimes is measured within a 100 meters radius. For simplicity we consider Euclidean distance (distance as the crow flies) around the place of transaction to draw a ring buffer. Figure 6 illustrates our method. We consider separately five types of infractions: burglaries, thefts without violence, thefts with violence, physical violence, and drug and immigration offenses.

The circle of 100 meters around the transactions on the edge of Paris may intersect out of Paris zones where crime data are not available. Therefore, we add in our estimations a dummy that indicates whether a transaction is located within 100 meters of the Parisian border.

Noise level: We measure the level of noise for each transaction based on the Bruitparif data. This regional observatory of noise developed a model to estimate noise levels by sources (road, rail and aircraft) for every location in Paris region. The model combines topographic data with very precise information concerning the noise sources (road surface, vehicle speeds, etc.). For each pair of geographical coordinates corresponding to one transaction, we have access to the noise level by source (in brackets of 5 dB) in building facades. The measures distinguish for each source day-time level and a weighted measure between daytime and nighttime noise. A dummy for the proximity (under 200 meters) to a garden and a dummy for the proximity to a calm zone (under 50 dB) are available as well.

[6]The ETAT 4001 dataset was geocoded for the 2008 year. The access to these data is governed by an agreement between GENES (*Groupe des Écoles Nationales d'Économie et de Statistiques*) and the INHESJ (*l'Institut National des Hautes Études de la Sécurité et de la Justice*). We thank Jean-Luc Besson for the geocoding and the use recommandations.

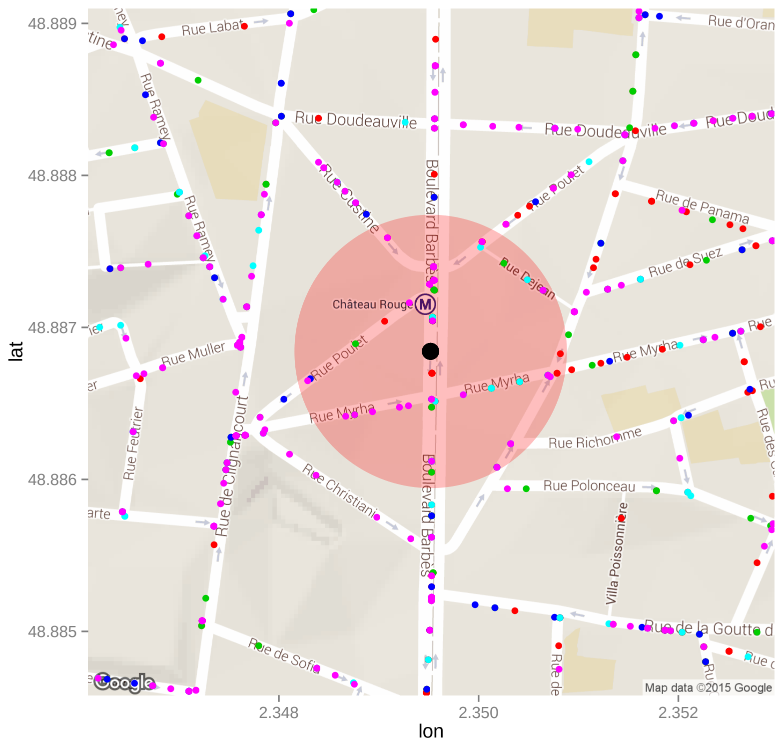


Figure 6: Counting Procedure

Source: ONDRP and authors' calculations.
Notes: The black cross represents a transaction. The transparent pink area is a 100 meter radius circle zone around the transaction. Blue points represent thefts with violence. Red points represent IRAS. Green points represent physical violence. Cyan points are burglaries and magenta points are thefts without violence.

Figure 7 shows a weighted measure of daytime and nighttime noise in Paris. The higher noise areas (more than 75dB) correspond the main road axes of the city and the ring road.

Type of street: Following Cheshire & Sheppard (1998), we exploit the type of streets as a proxy for the width of the street, and consequently for the view from the flat. Table II reports the frequencies of each type of location. More than 79% of the transactions are located in standard streets. Around 16% are on large streets, boulevards or avenues. Less than 1% are on quays.

School quality: The dwelling price is also potentially linked to the school quality. In France, parents are not free to choose the public school of their children. For public junior high school, catchment areas are defined. These residence-based assignment policy (the *Carte scolaire*) assigns one public junior high school-level to each pupil based on their precise

Figure 7: Weighted Measure between Daytime and Nighttime Noise

Notes: Noise levels are in dB.
Source: BruitParif

Table II: Type of streets

Type of streets	Number of transactions	%
Boulevard	8,553	15.7
Impasse	1,559	2.9
Square	683	1.3
Quay	410	0.8
Street	43,190	79.4
Total	54,395	100

Source: Bien database.

address. In Paris, there are 109 junior high school zones, that can be compared to the 992 IRIS zones. For each address of our transactions database, we use the public website of the Paris municipality to know the assigned public junior high school for the school year 2014-2015[7].

Furthermore, the statistical service of the French Ministry of Education provides us some data on public junior high schools in 2008: honours and success rates at the final exam (*Brevet des collèges*), repetition rate, and median scores. Figure 8 shows the success rate at the final exam for the different assignment areas in Paris.

The average success rate over the Parisian transactions of our sample is 71.49% with a standard error of 9.34. There is heterogeneity between arrondissements, some perform

[7]We assume that assignment areas have not changed since 2007 except for the creations of new junior high schools. In this case, we assign the closest junior high school to the transaction in 2007.

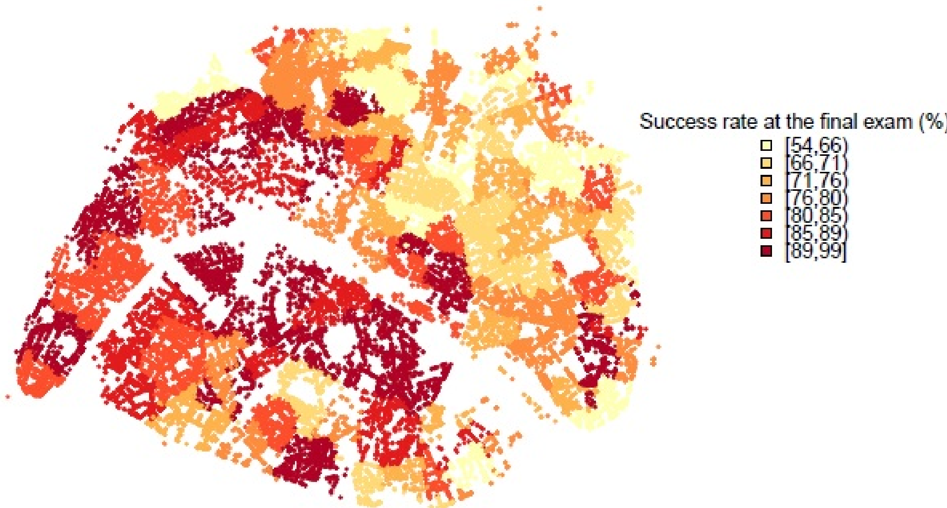


Figure 8: Success Rate at the Final Exam of the Public Junior High School Assigned to each Transaction

Source: Paris City Council, DEPP
Notes: each dot is a transaction

well above the Paris average and have homogeneous results within the arrondissement, such as the VI^{th} with an average success rate of 92% and a standard deviation of 4%, others have results below the Parisian average and more heterogeneity between schools of a same arrondissement, such as the XIX^{th}, with an average rate of 71% and a standard deviation of 10%. Table III provide further descriptive statistics on school quality by arrondissement.

III.4 Overview of Variables

Table IV reports the variability of the different variables at different administrative levels: within Paris (Total variability), within Grand-Quartier and within IRIS. Because the measure of the job accessibility is constructed at the modus-area level (corresponding more or less to groups of 9 IRIS), the within-IRIS variability of this variable is zero. The variability of the job accessibility variable within Grand-Quartier comes from both the fact that the Grand-Quartier areas are slightly larger than Modus areas, but also from the fact that the two delineations do not overlap. All the other variables are transaction-specific and the corresponding within IRIS variability is substantial.

When used as control variables in the regression, the presence of the variables described in table IV is indicated by the row "location characteristics".

Table III: Results at the year 9 examination of the assigned school

	Transactions	Success rate	Honours rate
75001	439	77.98 (10.13)	72.29 (7.13)
75002	740	78.96 (2.22)	65.82 (3.17)
75003	1225	71.41 (5.13)	65.62 (6.27)
75004	753	80.12 (10.3)	76.96 (4.31)
75005	1208	89.97 (6.01)	70.08 (11.56)
75006	1121	92.37 (4.32)	82.81 (3.32)
75007	1163	87.94 (3)	81.03 (4.91)
75008	894	94.78 (2.94)	74.23 (2.71)
75009	1852	89.83 (3.72)	76.53 (3.81)
75010	2667	81.37 (6.51)	73.27 (6.33)
75011	4771	69.79 (5.09)	70.08 (4.37)
75012	3093	71.76 (5.28)	69.55 (8.83)
75013	2905	76.27 (9.6)	72.37 (5.08)
75014	2814	78.73 (9.43)	75.73 (4.11)
75015	5920	81.24 (9.43)	72.52 (8.07)
75016	4248	82.14 (5.35)	72.96 (4.68)
75017	5323	84.34 (5.22)	77.23 (3.8)
75018	6242	81.51 (8.9)	74.07 (4.49)
75019	3288	70.97 (10.01)	68.15 (7.65)
75020	3729	69.07 (6.97)	69.08 (5.98)
Total	54395	71.49 (9.34)	67.76 (7.43)

Notes: In the first arrondissement, the average success rate at the Brevet in the assigned public junior high schools of the 439 observed transactions was 77.98% in 2007.
Source : DEPP, Paris municipality and authors' calculations

IV Results

IV.1 Explained Variance of Housing Prices

Before describing the results of hedonic prices, we compare the explained variance of housing prices between our different specifications. Table V reports the results.

Table IV: Variables overview

Variable	Definition	Scale	Mean	Total variability	Within Grand-Quartier	Within IRIS
Job accessibility	Employment potential	IRIS	13.56	0.16	0.06	.
Metro	Brackets of 100 meters	Transaction	2.87	1.22	1.13	0.77
Noise	Brackets of 5 dB	Transaction	3.10	1.32	1.28	1.18
Close to a green area	Dummy	Transaction	0.85	0.36	0.32	0.25
Burglaries	Number within 100 m	Transaction	6.26	4.13	3.61	3.00
Thefts without violence	Number within 100 m	Transaction	25.67	30.59	28.62	25.17
Thefts with violence	Number within 100 m	Transaction	13.28	12.42	0.11	9.29
Violences	Number within 100 m	Transaction	10.83	13.01	12.14	10.48
Drugs and immigration offenses	Number within 100 m	Transaction	8.53	15.50	14.47	11.94
School quality	Honors rate (%)	Transaction	72.29	7.13	4.84	2.42

Notes: We report the root mean squared errors of the regression of each variable on an intercept for the column "Total variability", on Grand-Quartier fixed effects for the next column and on the IRIS fixed effects for the last column.

Table V: Variance explained by the different groups of characteristics

Estimator	Without FE				Grand-Quartier FE	IRIS FE
	Surface	House Char.	Neigh. Char.	Location. Char.	Neigh. Char.	Neigh. Char.
Variables	log(price)	log(price)	log(price)	log(price)	log(price)	log(price)
	(1)	(2)	(3)	(4)	(5)	(6)
R^2	0.849	0.866	0.907	0.910	0.916	0.923
R^2 adj	0.849	0.866	0.907	0.910	0.916	0.922
Surface (polynomial)	Yes	Yes	Yes	Yes	Yes	Yes
Years and months	Yes	Yes	Yes	Yes	Yes	Yes
Dwelling characteristics	No	Yes	Yes	Yes	Yes	Yes
Neighbourhood characteristics	No	No	Yes	Yes	Yes	Yes
Number of observations	54395	54395	54395	53962	53962	53962

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Notes: Fixed effects at the Grand-Quartier level are introduced in the column (5) and fixed effects at the IRIS level are introduced in the column (6). The logarithm of the price is regressed only on the surface and on powers of the surface in column (1). We control for the characteristics of the house (number of rooms, floor, etc.) in column (2). We control for the neighbourhood characteristics (defined at the IRIS level) in column (3) and we control for the location characteristics (noise, crime,...) in column (4). We always introduce year and month dummies. Standard errors are clustered at the IRIS level.

For the estimations without fixed effects (columns 1 to 4), the explained variance ranges from 85% when we control for the surface (a 3-order polynomial) to 91.2% when we add all the covariates.

The explained variances for the two fixed-effects estimations are close to the one obtained with the estimation with the maximum number of controls. The characteristics we use to describe the real-estate property seem to explain particularly well the price dispersion.

IV.2 Hedonic Prices of the Housing Unit Intrinsic Characteristics

IV.2.1 Surface

The surface of the flat explains more than 74% of the variance of the logarithm of prices within a given year (by simple regression on the surface). But when we introduce powers of the surface, we explain up to 85% of the variance of the prices (see Table V). To capture these non-linearities, we apply a semi-parametric method *à la* Robinson.

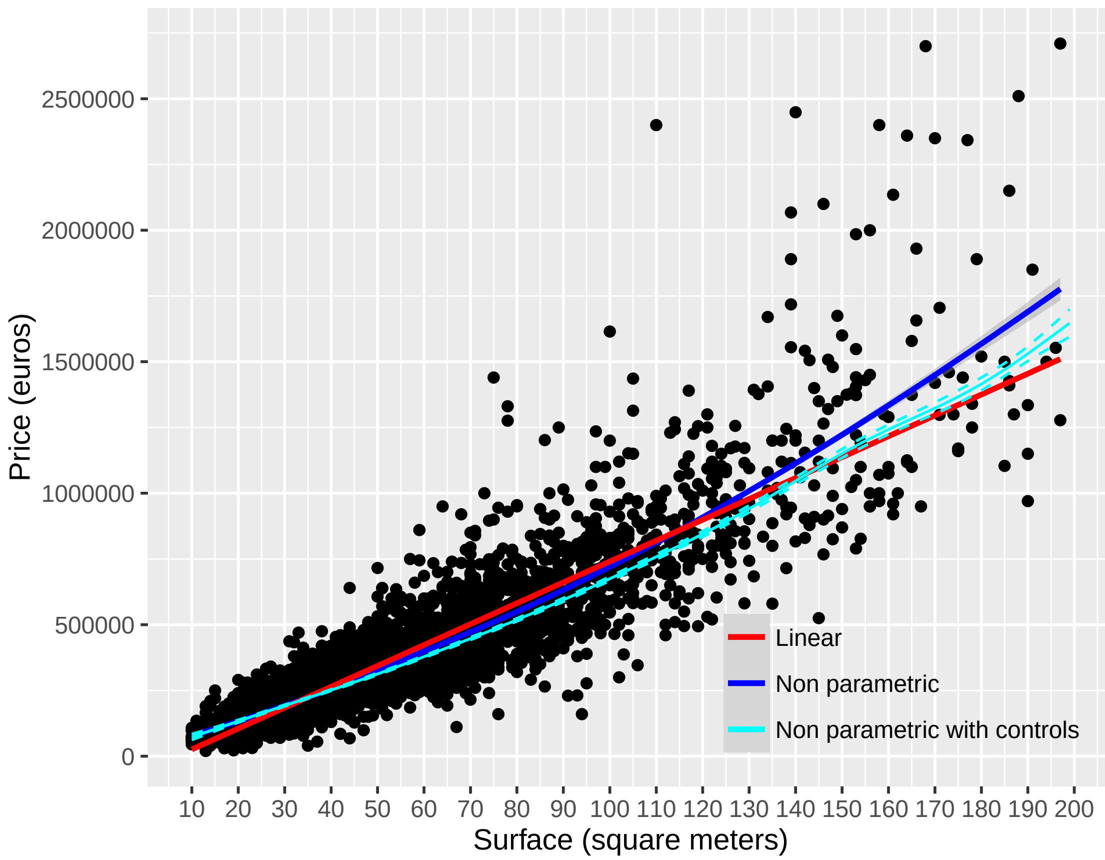


Figure 9: Hedonic Price of the Surface

Notes: The relationship between the price and the surface is estimated by a linear model, by a pure non-parametric method and by a non-parametric method with linear controls. This last method corresponds to a semi-parametric estimation using spline functions for the non-parametric variable and a linear part including all the attributes of the housing unit (among others the number of rooms) and the characteristics of the neighbourhood.

We estimate the relationship between the price and the surface (both in levels) with a sieve method (with spline functions) for a range of surfaces between 10 and 200 square meters. We represent on the figure 9 three estimations of this relationship. In red, we plot the estimated linear function, in blue the non-parametric relationship and in cyan the non parametric relationship when we control for all the other covariates (dwelling and location characteristics). The hedonic price function is always an increasing function of the surface. The three curves are close, meaning that the relationship is almost linear. In other terms, households who purchase smaller flats have the same marginal willingness to pay for an additional square meter than households who purchase larger flats.

IV.2.2 Other Attributes of the Dwelling

In Table VI we report the estimated hedonic prices for the dwelling characteristics controlling, or not, for the location characteristics and with, or without, fixed effects. Our results are consistent with the previous papers on hedonic prices of intrinsic characteristics using French data (see Maurer et al. 2004 for a comparison).

The estimated hedonic prices of the floor, of the occupancy status and of the presence of a lift are almost unchanged when neighbourhood characteristics or fixed effects are introduced. This means that these variables are uncorrelated with the location characteristics that influence the real-estate price. Thus, we find that the marginal willingness to pay to live on the first floor rather than on the ground floor, is at least equal to 5% of the value of the real-estate property (it is a lower bound, see II.2).

By contrast, the coefficients of the period of construction vary significantly between the regressions with (columns 2, 3 and 4) or without (column 1) location characteristics. We find a positive bias of the coefficient associated with the period of construction "before 1850" and a negative bias for the periods between 1914 and 1991. The dates of construction of the Parisian buildings are, indeed, not uniformly distributed in Paris, the oldest buildings are located in the center of Paris which is the most attractive zone. Nevertheless, even after the introduction of the characteristics of the neighbourhood, some periods of construction are more valued than others by the buyers. Thus, a lower bound of the difference of utilities between the purchase of a flat constructed during the period 1850-1914 and a flat constructed during 1914-1947 is equal to 0.9% of the value of the flat (IRIS fixed effects estimation). In other words, a buyer of a flat built between 1850 and 1914 is willing to pay at least 0.9% of the value of his flat to buy this flat rather than the same flat (other characteristics are considered as fixed) built between 1914 and 1947.

IV.3 Hedonic Prices of the Location Characteristics

Table VII presents the estimates of implicit prices of location characteristics for three specifications: without fixed effects, with Grand-Quartier fixed effects and with IRIS fixed effects

Table VI: Implicit prices of the Dwelling characteristics

Estimator	Without FE		Grand-Quartier FE	IRIS FE
Variables	log(price)	log(price)	log(price)	log(price)
	(1)	(2)	(3)	(4)
Basement	0.039	0.042	0.033	0.033
	(0.036)	(0.034)	(0.034)	(0.034)
Ground Floor	ref.	ref.	ref.	ref.
1^{st} floor	0.042***	0.049***	0.049***	0.050***
	(0.005)	(0.004)	(0.004)	(0.004)
2^{d} floor	0.072***	0.078***	0.077***	0.076***
	(0.005)	(0.004)	(0.004)	(0.004)
3^{d} floor	0.084***	0.090***	0.088***	0.088***
	(0.005)	(0.004)	(0.004)	(0.004)
4^{th} floor	0.095***	0.098***	0.096***	0.096***
	(0.006)	(0.005)	(0.005)	(0.005)
5^{th} floor	0.101***	0.104***	0.103***	0.104***
	(0.006)	(0.005)	(0.005)	(0.005)
6^{th} floor or higher	0.078***	0.089***	0.095***	0.104***
	(0.007)	(0.006)	(0.005)	(0.005)
1850 or before	0.202***	0.057***	0.017**	0.011*
	(0.016)	(0.009)	(0.007)	(0.006)
1850-1914	ref.	ref.	ref.	ref.
1914-1947	-0.017***	-0.009***	-0.008**	-0.009***
	(0.005)	(0.003)	(0.003)	(0.003)
1948-1969	-0.025***	-0.026***	-0.023***	-0.020***
	(0.006)	(0.004)	(0.004)	(0.003)
1970-1980	-0.050***	-0.011*	-0.003	0.008*
	(0.008)	(0.006)	(0.005)	(0.004)
1981-1991	-0.044***	-0.010	-0.005	-0.004
	(0.016)	(0.012)	(0.011)	(0.011)
1992-2000	0.063	0.076**	0.075***	0.056***
	(0.044)	(0.030)	(0.028)	(0.015)
2000-2010	0.076**	0.090***	0.093***	0.066**
	(0.037)	(0.032)	(0.030)	(0.026)
Lift	0.020***	0.015***	0.019***	0.021***
	(0.006)	(0.004)	(0.003)	(0.003)
Free	ref.	ref.	ref.	ref.
Partial occupation	-0.107***	-0.133***	-0.142***	-0.137***
	(0.035)	(0.030)	(0.029)	(0.027)
Occupied by the buyer	-0.166***	-0.171***	-0.169***	-0.164***
	(0.011)	(0.009)	(0.009)	(0.008)
Occupied by a tenant	-0.151***	-0.170***	-0.168***	-0.175***
	(0.019)	(0.013)	(0.013)	(0.008)
New	0.173***	0.214***	0.219***	0.199***
	(0.046)	(0.043)	(0.040)	(0.045)
Surface (polynomial)	Yes	Yes	Yes	Yes
Years and months dummies	Yes	Yes	Yes	Yes
Neighbourhood characteristics	No	Yes	Yes	Yes
Location characteristics	No	Yes	Yes	Yes
Number of observations	54395	53962	53962	53962
R^2	0.866	0.910	0.916	0.923
R^2 adj	0.866	0.910	0.916	0.922

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Notes: In the last two columns, fixed effects are introduced, at the Grand-Quartier level for the column 3 or at the IRIS level for the column 4. We always introduce year and month dummies. Standard errors are clustered at the IRIS level.

(see Figure 2 for the size of administrative districts, each arrondissement of Paris is divided in four *grands quartiers*). In every specification, we introduce all the intrinsic and extrinsic characteristics of the dwelling described in section 2.

Table VII: Implicit prices of the location characteristics

Estimator	Without Fixed Effects	Grand-quartier Fixed Effects	IRIS Fixed Effects
Variables	log(price) (1)	log(price) (2)	log(price) (3)
Metro			
Within 100 m of a metro station	-0.016**	-0.015**	-0.007
	(0.007)	(0.007)	(0.006)
Btw 100 and 200 m of a metro station	-0.008	-0.005	0.001
	(0.005)	(0.005)	(0.004)
Btw 200 and 300 m of a metro station	ref.	ref.	ref.
Btw 300 and 400 m of a metro station	0.008	0.009**	0.005
	(0.005)	(0.004)	(0.004)
Btw 400 and 500 m of a metro station	0.017**	0.013**	0.007
	(0.007)	(0.006)	(0.006)
More than 500 meters away from the metro	-0.005	-0.011	-0.013
	(0.012)	(0.010)	(0.012)
Job accessibility			
Employment potential (log)	0.254***	0.224***	(.)
	(0.027)	(0.037)	(.)
Crime			
Burglaries within a 100 m radius (*100)	-0.017	-0.053	-0.081**
	(0.051)	(0.040)	(0.038)
Thefts without violence within a 100 m radius (*100)	0.037***	0.007	0.000
	(0.011)	(0.009)	(0.008)
Thefts with violence within a 100 m radius (*100)	-0.049	0.016	0.037
	(0.044)	(0.035)	(0.028)
Violences within a 100 m radius (*100)	-0.059	-0.023	-0.025
	(0.041)	(0.031)	(0.026)
Drug and immigration offenses within a 100 m radius (*100)	-0.045**	-0.039***	-0.024*
	(0.018)	(0.013)	(0.013)
Within 100 m of the Parisian border	-0.054***	-0.046***	-0.030*
	(0.015)	(0.017)	(0.017)
Noise			
Less than 50 dB	ref.	ref.	ref.
Between 50dB(A) and 55dB(A)	-0.000	0.005	0.001
	(0.005)	(0.004)	(0.004)
Between 55dB(A) and 60dB(A)	-0.006	0.002	0.004
	(0.006)	(0.005)	(0.004)
Between 60dB(A) and 65dB(A)	-0.016***	-0.007	-0.006
	(0.006)	(0.005)	(0.004)
More than 65dB(A)	-0.042**	-0.032*	-0.009
	(0.019)	(0.017)	(0.015)
Close to a quiet area	0.010	0.011*	0.006
	(0.007)	(0.006)	(0.006)
Close to a green area	0.023***	0.003	0.002
	(0.006)	(0.005)	(0.005)
School quality			
Honours rate (normalized)	0.000	0.005	-0.002
	(0.004)	(0.003)	(0.004)
Honours rate (normalized) x Two rooms or more	0.010***	0.011***	0.011***
	(0.003)	(0.003)	(0.003)
Type of street			
Boulevard	0.004	0.005	-0.008
	(0.008)	(0.007)	(0.006)
Impasse	-0.014*	-0.013	-0.009
	(0.008)	(0.008)	(0.008)
Square	0.041***	0.047***	0.038***
	(0.014)	(0.014)	(0.013)
Street	ref.	ref.	ref.
Quay	0.102***	0.100***	0.106***
	(0.035)	(0.025)	(0.022)
Years and months dummies	Yes	Yes	Yes
Surface (polynomial)	Yes	Yes	Yes
Dwelling characteristics	Yes	Yes	Yes
Neighbourhood characteristics	Yes	Yes	Yes
Number of observations	53962	53962	53962
R^2	0.910	0.916	0.923
R^2 adj	0.910	0.916	0.922

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Notes: In the column 2, fixed effects at the Grand Quartier level are introduced. In the last column, fixed effects at the IRIS level are introduced. We always introduce year and dummies. Standard errors are clustered at the IRIS level.

Adding fixed effects has two consequences: the fixed effects control for unobservables and the location effects are identified based on intra-Iris variation. Consequently, the changes in the estimates can be explained either by the fact that unobservables are controlled for, or by the fact that little intra-zone variability remains. Even when we add IRIS fixed effects, some variability in unobservables may remain, hence some endogeneity.

Because we are interested in the least biased estimators, our preferred estimation is the one with IRIS fixed effects for most of the location variables. There are two exceptions: our preferred specification for job accessibility is the estimation with Grand Quartier fixed effects (job accessibility effect is not identified with IRIS fixed effects) and our preferred specification for school quality is the estimation on boundaries (stronger control of unobservables).

IV.3.1 Distance to Public Transport

The influence of the access to metro station is given by the first six rows of Table VII. We introduce dummy intervals of the same width (100 meters) to capture non linearities in the effect. We estimate five parameters and our reference value is between 200 and 300 meters of a metro station. Even if the coefficients are not precisely estimated, very close proximity to metro stations or relative remoteness from metro seem to be negatively valued by the buyers with an influence of around -1% of the housing price (insignificant for our preferred estimation, certainly due to weak remaining intra IRIS variability).

The influence of the distance to public transportation seems to be weak on the Parisian real estate prices in line with the fact that only 1% of the transaction are located more than 600 meters from public transportation (*i.e.* less than 10 minutes walk).

IV.3.2 Job Accessibility

Our preferred estimation (with Grand-Quartier fixed effects) shows that an increase of 1% of the employment potential leads to a change of 0.22% in housing prices. The specification is in log-log form, so the coefficient can be directly interpreted as an elasticity of 0.2. This coefficient is significant at the level of 1% (see Table VII). These results are relatively close to the previous findings for Berlin, Rogaland or London, where an accessibility elasticity of about 0.25-0.3 was indicated (see Ahlfeldt 2011, Osland & Thorsen 2008, and Ahlfeldt 2013).

IV.3.3 Crime

We break down infractions by category to compare the influence of each category on housing prices.

In our preferred specification, the marginal willingness to pay is negative and significant at least at 10% for burglaries and IRAS (Infractions revealed by police investigators: mainly drug and immigration offenses). An increase of one standard deviation (see Table IV) causes a decrease of around 0.3% of the housing prices for burglaries and of around 0.4% for IRAS. Coefficients for thefts, thefts without violence, and violence are not significant.

The significant and positive coefficient obtained for thefts with the regression without fixed effects is due to the positive correlation between the number of thefts and the quality of the neighbourhood (thefts are located around tourist areas and in the wealthiest zones).

In contrast with our results, Gibbons (2004) finds no measurable impact of burglaries on prices but estimates that a one-tenth standard deviation decrease in the local density of criminal damage adds 1% to the price of an average Inner London property. This last effect is much larger than the one we estimate.

Because crime data are not available outside Paris, we add in every specification a dummy that indicates whether a transaction is located within 100 meters of the Parisian border. This dummy is associated with a significant and negative effect of approximately 3%. This can be due either to a high level of crime in this zone, which would affect the prices negatively or to other unobserved variables.

IV.3.4 Noise

We use a weighted measure of nighttime noise and daytime noise at the front of the building to assess the influence of noise on prices. The noise levels correspond to five categories: from less than 50 dB to more than 65 dB. The reference category is set to "less than 50 dB".

In our preferred specification, the coefficients associated with the other levels of noise are insignificant, though negative for the two highest levels of noise.

When we do not introduce location dummies in the hedonic regression but we control for the dwelling and location characteristics, the levels of noise larger than 50 dB are negatively valued by the buyers. If we introduce fixed effects at the level of the Grand-Quartier, the coefficient for the highest level (more than 65 dB is equal to -3% and significant at 10%.

As already mentioned, two interpretations are possible: either the variability of the noise levels within the IRIS are too weak to identify the hedonic price of the noise or the noise level is correlated with some characteristics of the IRIS that influence housing prices. In this second case, because the magnitude of the coefficients is weaker with the fixed effects, this would mean that there exists a negative correlation between the noise level and the quality of the neighbourhood.

Boes & Nuesch (2011) find that aircraft noise reduces apartment rents by about 0.5% per decibel. Here if we consider the simple OLS estimate, the effect to be located in a zone where the average noise is 50 dB compared to a zone where the noise is 66 dB is estimated to be 4.3%, which is equivalent to 0.2% per decibel which is in line with the result of Boes & Nuesch (2011). However the effect of a noise measure in decibel on housing prices is certainly not constant and strongly depends on the level of noise.

For the dummies of proximity to a quiet or a green area, we find insignificant effect in our preferred specification. As for the noise coefficient, we can not distinguish between the absence of variability at the IRIS level and the positive correlation between the presence of a park and the quality of the neighbourhood.

IV.3.5 Type of Street

We try to measure the influence of the view by using the type of streets in which the transaction takes place. Last rows of Table VII reports the results. We find a strong and positive effect of the location on a quay or on a square, indicating an unobstructed view. The effect of being located on a quay (a 10% increase of the price compared to the reference category, the street) is stronger than the effect of being located on a square (a 4% increase of the price compared to the reference category, the street), certainly indicating a nicer view on the *Seine*.

IV.3.6 School Quality

Table VIII: Boundary discontinuity estimation for school

Estimator	Without FE		IRIS FE	OLS on boundaries	Boundary FE
Variables	log(price) (1)	log(price) (2)	log(price) (3)	log(price) (4)	log(price) (5)
Honours rate (normalized)	0.0630***	-0.0009	-0.0027	-0.0004	-0.0065*
	(0.0058)	(0.0040)	(0.0040)	(0.0044)	(0.0037)
Honours rate (normalized) x Two rooms or more	0.0107***	0.0108***	0.0115***	0.0123***	0.0141***
	(0.0038)	(0.0033)	(0.0031)	(0.0041)	(0.0040)
Surface (polynomial)	Yes	Yes	Yes	Yes	Yes
Years and months dummies	Yes	Yes	Yes	Yes	Yes
Characteristics of the house	Yes	Yes	Yes	Yes	Yes
Characteristics of the neighbourhood	No	Yes	Yes	Yes	Yes
Number of observations	54395	54395	54395	34276	34276
R^2	0.876	0.909	0.922	0.904	0.914
R^2 adj	0.876	0.908	0.921	0.904	0.913

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.010

Notes: The honour rate is normalized by the standard deviation. In the column 3, fixed effects at the IRIS level are introduced. In columns 4 and 5 the sample of dwellings is restricted to transactions that are located at less than 150 meters from a school boundary. We always introduce year and month dummies. Standard errors are clustered at the IRIS level for columns 1 to 3 and at the school attendance level for columns 4 and 5.

The school quality is proxied by the honours rate at the final exam of the assigned junior high school (year 9 examination called *Brevet des collèges*). Other measures were available: the success rate, the median marks in French, in Mathematics or in History for the same exam, or the repetition rate in year 9. But the honours rate turned out to be the variable with the highest explanation power (in terms of R^2). Because of high correlation between these dimensions, we choose to include only this variable in the regressions. We interact this variable with a dummy indicating that the dwelling is not a one-room apartment.

For this location attribute, our preferred estimation is the one with boundary fixed effects. We report in table VIII the results. In columns 4 and 5, we select transactions that are located at less than 150 meters from a school boundary: 34,000 transactions over 54,000 approximately. Our preferred specification in column 5 adds boundary fixed effects, i.e. fixed effects defined for groups of properties sharing the same school boundary segment. A one-standard deviation change in honours rate raises prices by around 1.2% for the dwellings with two rooms or more.

As highlighted by Black (1999), numerous papers overestimate the marginal willingness to pay for a better school because the authors do not take into account the positive correlation between the quality and the wealth of the neighbourhood and the quality of the school in the area. We show this positive link in the first column of Table VIII. We find a marginal willingness to pay of 7% of housing price for an increase of one standard deviation of the honours rate when we do not control for the characteristics of the neighbourhood.

Our preferred estimation gives slightly weaker effects than those found by Black (1999) or Gibbons et al. (2013): a change of one-standard deviation causes a 2.1% increase for the US and a 3% increase for the UK. They are closer to those found by Fack & Grenet (2010) on Parisian data: for the period 1997-2004 a standard deviation increase in public school performance raises housing prices by 1.4 to 2.4%. It should be noted that Fack & Grenet (2010) use the 2004 school data for a hedonic regression on dwellings purchased between 1997 and 2004 where we use the 2008 data for the period 2008 to 2010 avoiding a potential endogeneity bias.

More generally, our estimation seems to confirm that the capitalization of public school performance in the price of real estate in Paris is weaker than the one estimated on UK or US data (see Gibbons & Machin 2008 for a complete review). This could be due to the relatively weaker tuition fees in private schools in France than in the US or in the UK and a less developed private sector in the US: 8% of pupils in lower secondary school against 22% in France in 2012, see OECD (2014). Because private schools are more accessible in France, the capitalization of public school quality in housing prices could be weaker than elsewhere.

IV.4 Magnitude of the Effects

We try to provide a coherent comparison of the different effects in Table IX. We report in column 3 the effect in percentage of the housing value corresponding to the change in the variable X reported in column 2. We find substantial effects for surface, floor area, job accessibility, crime and school quality. By contrast, public transport accessibility and noise level seem to have weaker effects on price.

Table IX: Magnitude of the effects

Variable	ΔX	Effect in % of the price
Surface	+1%	1%
Floor area	2^{d} vs Ground floor	7.6%
Job Accessibility	+1 %	0.2%
Public transport accessibility	More than 500 m vs btw 200 and 300	-1.3%*
Noise	More than 65dB vs Less than 50dB	-0.9%*
Crime (drug and immigration offenses)	+1 sd	-0.4%
School quality	+1 sd	1.1%

Notes: The effects presented in this table are derived from the IRIS fixed effect estimation, except for job accessibility, for which the Grand Quartier fixed effect estimation is used. +1 sd corresponds to a positive change of one standard-deviation of the variable. Noise level is not continuous and therefore, standard deviation of this variable is not available. We report in column 3 the effect in percentage of the housing value corresponding to the change in X reported in column 2. Results with * are insignificant at 10%.

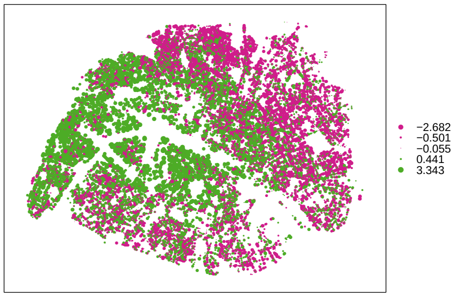
V Spatial Correlation

We find strong evidence of spatial autocorrelation in transaction prices. Figure 10 shows the spatial distribution of prices and residuals of the estimations in columns 2, 4 and 6 of Table V across Paris. Prices are represented by their deviation from the average. The larger the point, the higher the corresponding price or residual (in absolute value). Pink points represent positive deviations or positive residuals and green points represent negative ones.

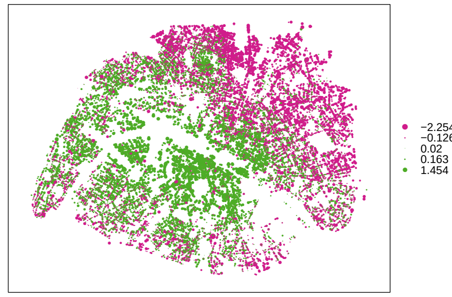
We find strong spatial correlation in the prices. Green points on the one hand, and pink points on the other hand, are geographically concentrated in the north-east for the greens and in the south-west and the centre of Paris for the pink. Moreover, the size of the points are relatively important. This concentration of pink and green points can also be seen for the residuals of the regression of the prices on the dwelling characteristics. However, the points are smaller, because a certain part of the price is explained by the housing unit intrinsic characteristics. On the contrary, there is no clear concentration of the same colour points when the characteristics of the location are also controlled for (corresponding to the fourth column of Table V) which means that the spatial correlation of the residuals is weaker in this case. And when we compare the spatial distribution of these residuals with the one of the residuals when we include IRIS fixed effects in the regression, we find a very similar picture, in line with the fact that the regression with the location characteristics strongly reduces the spatial correlation of the residuals.

To confirm these findings, we compute Moran's I or Mantel's test to check for spatial autocorrelation. Both test against the null hypothesis that there is no spatial autocorrelation. Moran's I is a parametric test which computes a correlation that is weighted by inverse distances. When fixed effects are added in the regression, autocorrelation can only arise within fixed effects. Therefore, in the inverse matrix distance, the cells between points in two different IRIS are set to zero.

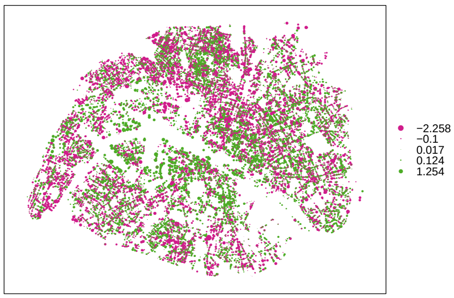
Mantel's test is semi-parametric, it examines the correlation between two distance matrices and generating a null distribution for this correlation by randomly permuting one of the matrices. Table X reports the results. We find that the spatial autocorrelation shrinks when we add variables in the regression and becomes insignificant at 10% for the regressions with dwelling characteristics and with dwelling and location characteristics for the Mantel's test and for the regression with IRIS fixed effects for Moran's and Mantel's tests.



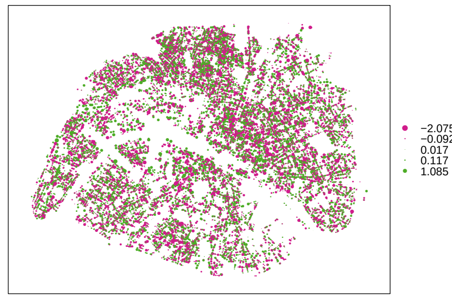
(a) Prices (log)



(b) Residuals (dwelling characteristics)



(c) Residuals (location characteristics)



(d) Residuals (IRIS FE)

Figure 10: Spatial Correlation

Notes: we represent the prices or the predicted residuals of the hedonic regression of the logarithm of the prices on different groups of covariates.

Table X: Spatial correlation indexes

	Prices	Residuals Dwelling char.	Residuals Location char.	Residuals IRIS FE
Moran index	21.97	27.85	6.03	0.90
p.value of Moran's test	2.90e-107	5.64e-171	8.44e-10	0.184
Mantel index	0.0346	0.00165	-0.0259	-0.0264
p.value of Mantel's test	0.0196	0.235	1	1
Dwelling characteristics	No	Yes	Yes	Yes
Location characteristics	No	No	Yes	Yes
Fixed effects	No	No	No	Yes

Notes: The p-values correspond to the test of the presence of spatial correlation in the data.

VI Conclusion

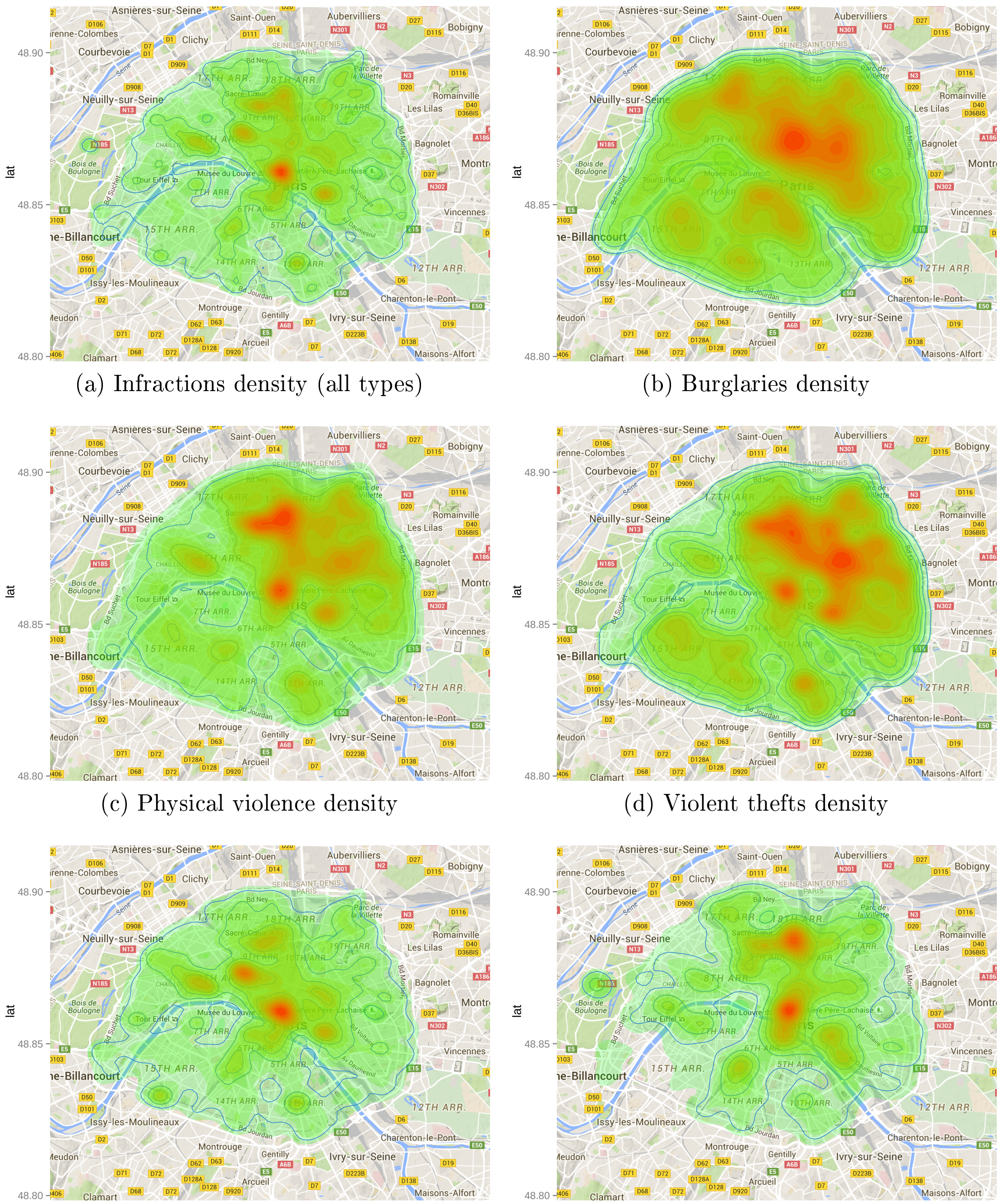
In this paper, first we develop a theoretical framework to show that hedonic prices can be interpreted as the marginal willingness to pay of the buyer at the equilibrium under assumptions of atomicity and quasilinearity of the utility function. Second, we present the estimations of hedonic prices for different attributes of the housing unit, intrinsic to the dwelling or related to its location. We construct a new database by adding to the housing transactions data, precise information about the neighbourhood of each real-estate property. We then compute the hedonic prices of the different attributes of the dwelling and its location. We run different estimations with and without geographical fixed effects, and with fixed effects at different geographical scales.

By adding local fixed effects, we control for zone-level unobservables and consequently we identify the effects only on intra zone variability. Even if we are able to introduce very local fixed effects, some intra variability on unobservables may remains. For instance, we do not control for the immediate proximity to social housing or for local urban renewal. These phenomena may influence housing prices and can be correlated with other location characteristics. In the case where an unbobserved price determinant is correlated with one of our characteristics within IRIS, an endogeneity bias may exist. If we think for instance that social housing is constructed close to metro stations, and that the proximity to social housing is negatively valued by buyers, we will under estimate the effect of the distance to metro on prices. But we think that conditional on the IRIS location, the precise location of these local transformations are rather random and depends on local opportunities, such as building availability.

To sum up, we find a substantial positive marginal willingness to pay for job accessibility and school quality and negative marginal willingness to pay for crime rate in the area.

As emphasized in the introduction, the hedonic method gives an estimate of the marginal willingness to pay of the buyers at the equilibrium, and is conditional on buyer's characteristics. A complete investigation of hedonic prices would therefore require to investigate the heterogeneity of this marginal willingness to pay. This is beyond the scope of this paper but should deserve attention in future research.

A Crime Maps



(a) Infractions density (all types)

(b) Burglaries density

(c) Physical violence density

(d) Violent thefts density

(e) Thefts without violence density

(f) Drug and immigration offenses

Figure 11: Crime Densities

Source: ONDRP and authors' calculations.

On Figure 11, we represent spatial densities smoothed over space of infractions. We first represent the density of all types of infractions. The right bank of Paris has a higher density of crimes and we observe some peaks for particular places: around the *Champs-Elysées* Avenue, in the proximity to *Châtelet-les-Halles* and a high density of infractions in the XVIIIth. Since the different types of infractions could have different impacts on the neighbourhood reputation depending on the risk perceived by the residents, we follow the nomenclature of the ONDRP to construct some broad categories of infractions: burglaries, violent thefts, thefts without violence, violence and infractions revealed by the police activity (mainly drug and immigration offenses).

We then represent the spatial distribution of burglaries in Paris. This type of infraction concerns almost equally the different neighbourhoods of Paris.

For the spatial distribution of physical violence infractions, we observe a very high concentration of this type of infraction around the neighbourhood of *la Goutte d'Or* and more generally a high concentration in the North East. We note also a peak around *Châtelet-les-Halles*.

The map with the spatial distribution of violent thefts is roughly similar to the map of the physical violence density, but a new area with a very high concentration appears: the neighbourhood of *Belleville*.

The thefts without violence are less diffused than the violent events. We see a high concentration at the centre of Paris and some particularly exposed areas, mostly tourist areas, such as the neighbourhoods of *Opéra*, of *Bastille*, the *Champs-Elysées* Avenue and the *Sacré-Cœur*.

Last, the drug and immigration offenses are concentrated only in certain areas in line with the fact that police controls take place in the zones that are more likely to be concerned by these infringements.

B Neighbourhood Characteristics

Table XI: Overview of neighborhood characteristics

Variable	Mean	Total variability	Within Grand-Quartier
Public amenities			
Number of supermarkets	0.57	0.84	0.79
Number of local stores	6.45	4.91	4.42
Number of retail shops	18.12	20.09	13.08
Number of local services	20.22	17.05	10.19
Number of restaurants	20.16	17.56	11.13
Census Variables			
Proportion of houses	0.01	0.02	0.02
Proportion of studio apartments	0.24	0.08	0.07
Proportion of homeowners	0.36	0.10	0.09
Proportion of social housing	0.14	0.19	0.16
Unemployment rate	0.11	0.03	0.02
Proportion of students	0.13	0.04	0.03
Proportion of non-graduated	0.13	0.06	0.05
Proportion of CAP-BEP	0.08	0.03	0.02
Proportion of CEP-BEPC	0.10	0.03	0.02
Proportion of BAC	0.15	0.03	0.02
Proportion of BAC+2	0.12	0.02	0.02
Proportion of more than BAC+3	0.41	0.11	0.06
Proportion of 18-24 year old	0.11	0.03	0.02
Proportion of 25-39 year old	0.29	0.06	0.04
Proportion of 40-64 year old	0.30	0.03	0.03
Proportion of more than 65 year old	0.14	0.05	0.03
Proportion of immigrants	0.20	0.06	0.05
Fiscal data			
Median fiscal income	9525.16	3635.17	2088.59

Notes: We report the root mean squared errors of the regression of each variable on an intercept for the column "Total variability", on Grand-Quartier fixed effects for the column "Within Grand-Quartier".
Source: *Base Permanent des Équipements*, Census, ERFS

References

Ahlfeldt, G. (2011), 'If alonso was right: Modeling accessibility and explaining the residential land gradient', *Journal of Regional Science* **51**(2), 318–338.

Ahlfeldt, G. M. (2013), 'If we build it, will they pay? predicting property price effects of transport innovations', *Environment and Planning A* **45**(8), 1977–1994.

Ahlfeldt, G. M. & Holman, N. (2016), 'Distinctively different: A new approach to valuing architectural amenities', *The Economic Journal* pp. n/a–n/a.

Ahlfeldt, G. M. & Wendland, N. (2016), 'The spatial decay in commuting probabilities: Employment potential vs. commuting gravity', *Economics Letters* **143**(C), 125–129.

Bajari, P. & Benkard, C. L. (2005), 'Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach', *Journal of Political Economy* **113**(6), 1239–1276.

Bartik, T. J. (1987), 'Estimating Hedonic Demand Parameters with Single Market Data: The Problems Caused by Unobserved Tastes', *The Review of Economics and Statistics* **69**(1), 178–80.

Bayer, P., Ferreira, F. & McMillan, R. (2007), 'A unified framework for measuring preferences for schools and neighborhoods', *Journal of Political Economy* **115**(4), 588–638.

Berry, S., Levinsohn, J. & Pakes, A. (1995), 'Automobile Prices in Market Equilibrium', *Econometrica* **63**(4), 841–90.

Bidoux, P.-E., Caenen, Y. & Trigano, L. (2017), 'Déplacements domicile-travail. à paris, le vélo est dépassé par le métro', *Insee Analyses Ile-de-France* **16**.

Black, S. E. (1999), 'Do Better Schools Matter? Parental Valuation Of Elementary Education', *The Quarterly Journal of Economics* **114**(2), 577–599.

Boes, S. & Nuesch, S. (2011), 'Quasi-experimental evidence on the effect of aircraft noise on apartment rents', *Journal of Urban Economics* **69**(2), 196–204.

Brown, J. N. & Rosen, H. S. (1982), 'On the estimation of structural hedonic price models', *Econometrica* **50**(3), pp. 765–768.

Can, A. (1990), 'The measurement of neighborhood dynamics in urban house prices', *Economic Geography* **66**(3), 254–272.

Cavailhès, J. (2005), 'Le prix des attributs du logement', *Economie et statistique* **381**(1), 91–123.

Chay, K. Y. & Greenstone, M. (2005), 'Does Air Quality Matter? Evidence from the Housing Market', *Journal of Political Economy* **113**(2), 376–424.

Chen, X. (2007), Chapter 76 large sample sieve estimation of semi-nonparametric models, Vol. 6, Part B of *Handbook of Econometrics*, Elsevier, pp. 5549 – 5632.

Cheshire, P. & Sheppard, S. (1998), 'Estimating the demand for housing, land, and neighbourhood characteristics', *Oxford Bulletin of Economics and Statistics* **60**(3), 357–382.

Court, A. T. (1939), 'Hedonic Price Indexes with Automotive Examples', *The Dynamics of Automobile Demand.* .

Diewert, E. (2003), 'Hedonic regressions: A review of some unresolved issues'.

Ekeland, I., Heckman, J. J. & Nesheim, L. (2004), 'Identification and Estimation of Hedonic Models', *Journal of Political Economy* **112**(S1), S60–S109.

Fack, G. & Grenet, J. (2010), 'When do better schools raise housing prices? evidence from paris public and private schools', *Journal of Public Economics* **94**(1–2), 59 – 77.

Gibbons, S. (2004), 'The Costs of Urban Property Crime', *Economic Journal* **114**(499), 441–463.

Gibbons, S. & Machin, S. (2003), 'Valuing English primary schools', *Journal of Urban Economics* **53**(2), 197–219.

Gibbons, S. & Machin, S. (2005), 'Valuing rail access using transport innovations', *Journal of Urban Economics* **57**(1), 148–169.

Gibbons, S. & Machin, S. (2008), 'Valuing school quality, better transport, and lower crime: evidence from house prices', *Oxford Review of Economic Policy* **24**(1), 99–119.

Gibbons, S., Machin, S. & Silva, O. (2013), 'Valuing school quality using boundary discontinuities', *Journal of Urban Economics* **75**, 15 – 28.

Goujard, A. (2011), 'The externalities from social housing, evidence from housing prices'. Job Market Paper :1-22.

Heckman, J. J., Matzkin, R. L. & Nesheim, L. (2010), 'Nonparametric Identification and Estimation of Nonadditive Hedonic Models', *Econometrica* **78**(5), 1569–1591.

Hulten, C. R. (2003), ‘Price hedonics: a critical review’, *Economic Policy Review* (Sep), 5–15.

Koster, H. R. A., van Ommeren, J. N. & Rietveld, P. (2016), ‘Historic amenities, income and sorting of households’, *Journal of Economic Geography* **16**(1), 203–236.

Lancaster, K. J. (1966), ‘A New Approach to Consumer Theory’, *Journal of Political Economy* **74**, 132.

Lynch, A. K. & Rasmussen, D. W. (2001), ‘Measuring the impact of crime on house prices’, *Applied Economics* **33**(15), 1981–1989.

Machin, S. (2011), Houses and Schools: Valuation of School Quality through then Housing Market - EALE 2010 Presidential Address, CEP Occasional Papers 29, Centre for Economic Performance, LSE.

Maurer, R., Pitzer, M. & Sebastian, S. (2004), ‘Hedonic price indices for the paris housing market’, *Allgemeines Statistisches Archiv* **88**(3), 303–326.

Nelson, J. P. (2004), ‘Meta-Analysis of Airport Noise and Hedonic Property Values’, *Journal of Transport Economics and Policy* **38**(1), 1–27.

Nesheim, L. (2006), Hedonic price functions, CeMMAP working papers CWP18/06, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Nesheim, L. (2008), hedonic prices, *in* S. N. Durlauf & L. E. Blume, eds, ‘The New Palgrave Dictionary of Economics’, Palgrave Macmillan, Basingstoke.

Ngai, L. R. & Tenreyro, S. (2014), ‘Hot and cold seasons in the housing market’, *The American Economic Review* **104**(12), 3991–4026.

OECD (2014), ‘Education at a glance 2014’.

Osland, L. & Thorsen, I. (2008), ‘Effects on housing prices of urban attraction and labor-market accessibility’, *Environment and Planning A* **40**(10), 2490–2509.

Pakes, A. (2003), ‘A Reconsideration of Hedonic Price Indexes with an Application to PC’s’, *American Economic Review* **93**(5), 1578–1596.

Robinson, P. M. (1988), ‘Root- N-Consistent Semiparametric Regression’, *Econometrica* **56**(4), 931–54.

Rosen, S. (1974), ‘Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition’, *Journal of Political Economy* **82**(1), 34–55.

Smith, V. K. & Huang, J.-C. (1995), 'Can Markets Value Air Quality? A Meta-analysis of Hedonic Property Value Models', *Journal of Political Economy* **103**(1), 209–27.

Trannoy, A., Michelangeli, A. & Barthélémy, F. (2007), 'La rénovation de la Goutte d'Or est-elle un succès ? Un diagnostic à l'aide d'indices de prix immobilier', *Économie et Prévision* **180**(4), 107–126.

Triplett, J. (2004), Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products, OECD Science, Technology and Industry Working Papers 2004/9, OECD Publishing.

Yinger, J. & Nguyen-Hoang, P. (2016), 'Hedonic vices: Fixing inferences about willingness to pay in recent house-value studies', *Journal of Benefit-Cost Analysis* **7**(2), 248–291.