The relationship between institutional investors and house prices:

The case of Paris

August 20th, 2021

H.J. van Haften - S4185269

Msc Real Estate Studies

University of Groningen

Colofon

Title	The relationship between institutional investors and house prices: *The case of Paris*
Version	1.3
Author	H.J. van Haften
Supervisor	Prof. dr. Ir. A.J. van der Vlist
Assessor	Prof. dr. E.F. Nozeman
E-mail	h.j.van.haften@student.rug.nl
Date	20 August 2021

Abstract

Paris is one of the most expensive cities to live in, with prices still rising. Investors may be part of the cause, as the growing presence of investors results in additional demand for housing. This study aims to investigate the relationship between institutional investments in residential real estate on house prices in the area. We use a unique approach of measuring investment activity within a radius of each house transaction and considering investment values rather than number of transactions. We find a positive correlation between house prices and residential investments in the close vicinity. Furthermore, we find heterogeneity across house types and significant differences between the four geographical departments of the Greater Paris region that are included in the analysis.

1. Introduction

1.1 Motivation

Paris has been one of the most expensive cities to live in (The Economist, 2020). In the first half of 2020, the average price per square meter of an apartment in Paris was with 10,175 euros per month the fourth highest of Europe (Statista, 2020). Despite the recession of the French economy, house prices have kept rising (Insee, 2021a). House prices in Paris specifically have had a significantly greater increase compared to the rest of the country (Delmendo, 2020; Insee, 2021a). Between early 2010 and mid 2012, house prices rose with 27 percent in Paris, in comparison to a country average of 8 percent. After 2012, house prices in Paris kept rising and at a higher rate than house prices in France (Ministère de la Transition Écologique, 2021). Space constraints and high population density make it difficult to construct new housing in Paris (Baleó, 2019). Therefore, the construction of new houses is staying behind, causing a lack of residential real estate for the current demand (AEW, 2020; Laurent, 2020).

One of the reasons of the rising house prices in Paris is the persistently high demand for housing, that originates from owner-occupiers and investors. Investors may be part of the cause of rising house prices in Paris, as these current conditions of the Parisian residential market result in a strong housing demand from investors (Savills, 2020; AEW, 2020; Bourla, 2021). The growing presence of investors in the housing market results in additional demand for homes on the buyer's market. Based on market fundamentals, increasing demand while supply remains equal results in higher prices (Bourla, 2021; Strangler, 2020; Laurent, 2020).

For Parisian policy makers, increasing knowledge about the drivers for house prices can be beneficial for the assessment of current policies and government intervention measures. Furthermore, institutional investors require reliable information about the development of house prices as well to base their investment decisions on. Therefore, this study aims to explore the relationship between institutional investors and house prices in Paris.

1.2 Literature review

Several studies investigate the relationship between investor activity and house prices, of which most are conducted in the United States. Allen et al. (2017), Smith & Liu (2020) and Mills et al. (2019) estimate house prices by including a dummy or categorical variable for purchases made by investors. Cvijanoviç & Spaenjers (2020) study the housing market in Paris, with a focus on foreign buyers instead of institutional investors. Alternatively, Lambie-Hanson et al. (2015) investigate the relationship between investors and house prices with a qualitative approach. Lambie-Hanson et al. (2019) study the relationship between net share of institutional investors and house price growth. Literature about institutional investors and house prices in the Parisian housing market has not been found. Studies of Paris focus either on hedonic models (Maurer et al. 2004; Baltagi et al., 2015), or on buyer types other than (institutional) investors (Cvijanoviç & Spaenjers, 2020)

First, part of the literature focuses on estimating house price levels with investor buyers or investor activity as key explanatory variable. Allen et al. (2017) investigate the relationship between different types of investors and house prices in distressed housing markets specifically. The dataset contains sales of single-family homes in Miami, United States in the period of 2009 to 2013. They compare sales prices of houses bought by investors with houses bought by single-purchase buyers. Investors are divided into four categories based on their size, with institutional investors being one of them. Results indicate that institutional investors purchased at an average discount of 7.7 percent compared to single-purchase buyers and that other groups of investors buy at greater discounts. Moreover, results show an increase of 0.20 percent in house prices in a market where 10 percent of the houses is purchased by investors. Allen et al. (2017) conclude that investors are likely to be responsible for the overall price recovery in Miami, because of their higher buyer power. Furthermore, they conclude that properties purchased by investors drive up prices in distressed markets. Smith & Liu (2020) study differences in purchase prices for single-family detached houses between institutional investors and owner-occupier buyers in Atlanta and Georgia, United States. In their study with data between 2000 and 2014, they find a similar result of institutional investors buying for a 6.3 to 11.8 percent lower price than owner-occupiers in the time after the real estate crisis (Smith & Liu, 2020). Mills et al. (2019) study transaction data of single-family homes sold between 2000 and 2014. Several regressions are conducted with share of single-family homes purchased by investors as dependent variable. More relevant for this thesis is the OLS regression where they estimate the log price per square foot paid by investors with a function of investor types and the average characteristics of the properties purchased by each investor type. Their evidence implies that buy-to-rent investors contributed to an increasing housing demand between 2012 and 2014, with slightly higher house prices and marginally lower vacancy rates as a result. However, this study does not distinguish institutional investors as separate group. Alternatively, Cvijanoviç & Spaenjers (2020) examine the effect of out-of-country buyers on the housing market in Paris between 1992 and 2016. However, only transactions between households are included in the analysis, excluding all other foreign non-resident buyers such as institutional investors. The data is used in a linear regression, measuring the effects of out-of-country buyers on real estate prices in Paris. The regressor in the equation is the cumulative inflow of non-resident and resident foreigners combined. Conditional correlations between the inflow of foreigners and price trends are estimated, with a significant positive correlation as a result (Cvijanoviç & Spaenjers, 2020). This positive correlation can be explained with rising demand and the market fundamentals as well, but the study focuses on buyer types other than institutional investors. Lambie-Hanson et al. (2015) analyse investor activity in purchasing foreclosed properties, with data of

4,700 residential property sales in Suffolk County, United States, from the years 2007 to 2012. Conclusions about how investors have affected local markets are based on a qualitative analysis. Accordingly, they conclude that investors helped absorbing a major part of the distressed properties. However, in addition to the fact that this study has a qualitative instead of quantitative approach, the consequences of a market without institutional investors buying foreclosed properties are unclear (Lambie-Hanson et al., 2015).

Second, literature is available about house price growth rates in relation to investor activity. Lambie-Hanson et al. (2019) study 11.8 million transactions of single-family houses in the United States that were purchased by institutional investors between 2000 and 2014, a period that contains different cyclical stages of the housing market. These regressions are conducted on zip code level, with house price growth rate as dependent variable and net share of institutional investors as key-independent variable. They find a significant positive relationship between the two variables, indicating that an increase in the net share of institutional investors results in a higher house price growth rate (Lambie-Hanson et al., 2019). It should be noted that the dependent variable is different from the dependent variable used in this thesis, which is house price level instead of house price growth rate.

Several papers study the relationship between investor activity and house prices. However, none of these studies relate to Paris. Allen et al. (2017) and Lambie-Hanson et al. (2015) only consider effects in distressed markets or from foreclosed properties in Miami and Suffolk County. The studies from Lambie-Hanson et al. (2019) and Mills et al. (2019) include a longer time period and different cyclical stages of the housing market, but focus solely on single-family houses in the United States. Studies that do focus on Paris are either general hedonic studies or focus on foreign buyers. Considering that the Paris housing market is not similar to the general housing market of the United States and the fact that institutional investors often buy multi-family homes instead of single-family homes, it is relevant to study the influence of institutional investor activity on house prices in Paris.

1.3 Research problem statement

The aim of this study is to gain insight in the relationship between real estate investments of institutional investors and house prices in Paris. Accordingly, results of the study will contribute to the knowledge available about institutional investors in the housing market and the determination of house prices. The following research question will be attempted to answer: *"To what extent are residential real estate investments of institutional investors associated with higher house prices in Paris?"*

The answer to the central question will be based on three sub-questions:

1. What is the theoretical relationship between institutional investors and house prices?
2. What is the quantitative relationship between institutional investments and house prices in Paris?
3. What are the differences across house types and geographical departments?

What is the theoretical relationship between institutional investors and house prices?

The first question serves to gain insight in the theoretical background of the relationship between investor activity and house prices. Therefore, we review academic studies of house value drivers and studies on the relationship between investors and house prices. With these studies, we determine important house value drivers for our hedonic model and we formulate the hypotheses based on their findings. Based on the theoretical framework, we develop the following conceptual model:

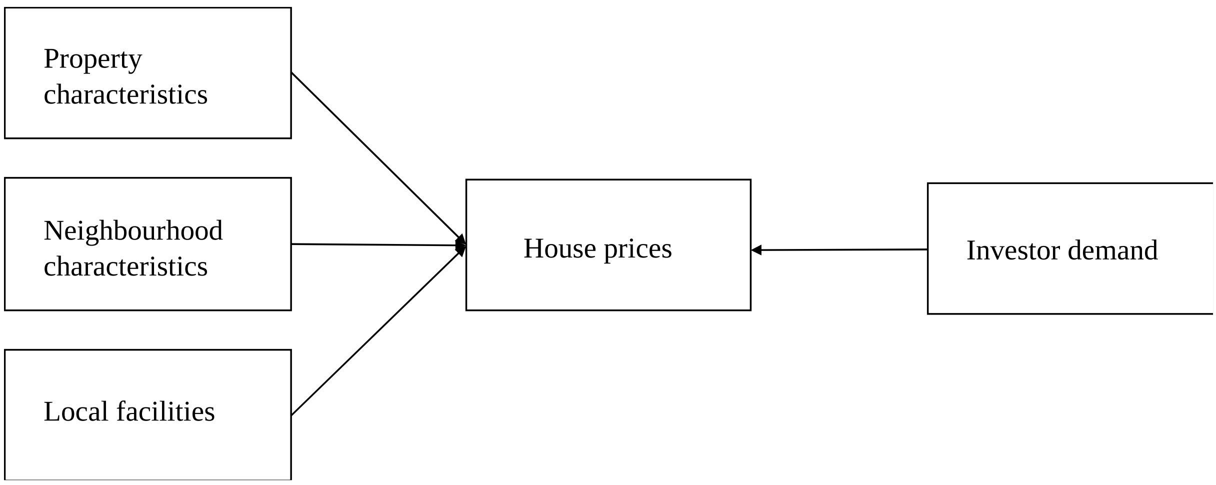

Figure 1. Conceptual model

What is the quantitative relationship between institutional investments and house prices in Paris?

With the second sub-question, we will explore the quantitative relationship between house prices and institutional investments in housing. By cooperating with the global investment manager AEW, we are able to analyse the Real Capital Analytics investment transactions database and connect it to public data sources from Demande de Valeurs Foncières (DVF), RATP, Mairie des Paris and Grand Paris Seine Ouest. We use the transaction data from the period of 2014 to 2020. The housing transactions data from DVF is the basis for the analysis, with information on the prices, locations and characteristics of the houses. Then, we calculate distances to the closest public transport stops and closest parks with coordinates. Subsequently, we calculate total investment values within a radius of 1500, 2000 and 2500 metres of each housing transaction to use as our key explanatory variables. With this data, we estimate a hedonic model with an Ordinary Least Squares regression.

What are the differences across house types and departments?

We use the third research question to test heterogeneity across the observations. As indicated by AEW, the database with commercial housing transactions consists mainly of multi-family properties. Therefore, it is interesting to discover whether these investments have a different impact on multi-family homes than on single-family homes. Different types of houses can have a different group of buyers, so investors are possibly competing more with buyers from the condominium market than with the demand for single-family homes. Therefore, we will explore if associations between investments and house prices differ across housing types. Furthermore, we test heterogeneity across the four geographical departments included in the study area. Considering that the departments have different levels of investor activity and vary in average square meter price, they could be defined as different housing market segments. Therefore, there is a possible difference in the effect of the investment levels on house prices, which will be assessed. We test heterogeneity with Chow tests.

1.4 Reading guide

This chapter is followed by the theoretical framework, which creates the theoretical basis of the research and includes the formulated hypotheses. Second, we describe the data and the applied method in chapter 3, containing detailed information about the variables and models. The fourth chapter includes the empirical results and explanations. The report is completed with a conclusion and discussion. Finally, additional information is included in the appendices.

2. Theory

2.1 House value drivers

There are several drivers that are assumed to have an influence on house prices. According to Rosen (1974), house prices are the sum of shadow prices for individual property characteristics. Therefore, house prices depend on the property characteristics and the way these property characteristics are perceived by house buyers. The values that are assigned to these houses, depend on the willingness to pay for these characteristics. According to Harding et al. (2003), housing supply is assumed to be sufficiently large to be able to identify the values for each housing characteristic by comparing houses that differ in only a single attribute. As a result, the shadow prices of housing characteristics are well defined and bargaining power does not affect house prices. Accordingly, house prices are a result of the housing characteristics combined with the sum of the shadow prices for each characteristic. Given the fact that the supply cannot immediately adjust to a change in demand, studies often assume a fixed supply in their house price estimations. In this study, we only include variables related to the demand side of the housing market. Accordingly, the hedonic regression estimates the value of a house, based on the individual contributions of each characteristic to this value. The property characteristics can be classified into three categories (Dubin, 1992): property characteristics, local amenities and neighbourhood characteristics.

The majority of the property features included in the hedonic models are size and quality related. First, size refers to indoor or outdoor space. Indoor space includes surface area of the house, number of rooms and lot size in the cases of single-family homes. (Allen et al., 2017; Maurer et al., 2004; Baltagi et al., 2015; Cvijanovic & Spaenjers, 2015; Gouriéroux & Laferrère, 2009; Dubin, 1992; Bresson & Hsiao, 2011; Dubé & Legros, 2014; Halvorsen & Pollakowski, 1981). Number of bedrooms and bathrooms are used as well (Allen et al., 2017; Mills et al., 2019; Baltagi et al., 2015; Violand & Simon, 2007; Gouriéroux & Laferrère, 2009; Dubin, 1992; Bresson & Hsiao, 2011; Dubé & Legros, 2014; Halvorsen & Pollakowski, 1981). Baltagi et al. (2015) and Bresson & Hsiao (2011) added a dummy for the presence of a maid's room in their studies about Parisian property prices between 1990 and 2003. Furthermore, the presence of a garage, basement or attic is included in several studies (Maurer et al., 2014; Baltagi et al., 2015; Dubin, 1992; Dubé & Legros, 2014). Outdoor space is included in various forms, whereas the presence of a terrace, balcony or garden is found to have an influence on house prices (Maurer et al., 2004; Baltagi et al., 2015; Cvijanovic & Spaenjers, 2015; Violand & Simon, 2007; Gouriéroux & Laferrère, 2009; Dubin, 1992; Bresson & Hsiao, 2011; Dubé & Legros, 2014). Logically, the aforementioned size related variables have a positive impact on house prices. Second, house prices have been related to house quality in previous studies. Amédée-Manesme et al. (2017) distinguish duplex and triplex houses in their study about apartments in Paris and Dubin (1992) found a significant difference between detached and non-detached units. Additionally, Cvijanovic & Spaenjers (2015), Bresson & Hsiao (2011), Maurer et al. (2004), Baltagi et al. (2015), Amédée-Manesme et al. (2017) and Dubé & Legros (2014) use floor level as indicator for house prices. However, for both house types and floor levels, the different studies have conflicting results. In numerous studies, property amenities are added to the regression as well. Allen et al. (2017) includes a dummy for having a pool in their study about Miami and Dubin (1992) uses the presence of a fireplace and air conditioning as control variables in his study about Baltimore. These variables are not present in any of the French studies. However, Gouriéroux & Laferrère (2009) do include the presence of an elevator in their study on France, as well as several Parisian hedonic studies do (Maurer et al., 2004; Cvijanovic & Spaenjers, 2015; Dubé & Legros, 2014). Maurer et al. (2004) and Dubé & Legros (2014) display the regression results of having an elevator on house prices, which is a positive significant effect in both cases. Lastly, Cvijanovic & Spaenjers (2015), Dubin (1992) and Bresson & Hsiao (2011) included a variable related to parking places and display a positive relationship between parking places and house prices. Furthermore, the age of the construction is used frequently to estimate house prices, either as continuous variable or as categories of construction periods (Allen et al., 2017; Maurer et al., 2004; Mills et al., 2019; Baltagi et al., 2015; Cvijanovic & Spaenjers, 2015; Violand & Simon, 2007; Gouriéroux & Laferrère, 2009; Halvorsen & Pollakowski, 1981; Bresson & Hsiao, 2011; Dubé & Legros, 2014). Allen et al. (2017) and Mills et al. (2019) include the age as continuous variable and both find a significant negative effect on house prices. However, Maurer et al. (2004), Baltagi et al. (2015), Violand & Simon (2007) and Bression & Hsiao (2011) find varying results in both significance and sign of the coefficients between the different age categories they include. Dubé & Legros (2014) do find all significant negative results, where construction periods are compared to the reference category of houses built before 1850. The largest negative coefficients are from houses that are built between 1948 and 1980. The smallest negative effect is from the most recent construction period, which is between 1992 and 2000 (Dubé & Legros, 2014).

Local amenities are facilities that are available within a certain distance. These local amenities can be part of the neighbourhood, but can be within the administrative borders of the neighbourhood as well. Therefore, local amenities can influence the neighbourhood characteristics, but can still be distinguished as a separate category. According to Brueckner et al. (1997), urban amenities can be divided into three categories: natural amenities, historical amenities and modern amenities. Paris differentiates itself as a city because of its amenity advantage over the suburbs, because if its historical architecture and attractive scenery. Brueckner et al. (1997) also claim that the Seine river in Paris is an important natural amenity for residents. Palma et al. (2007) have conducted an empirical examination of local amenities in the Paris metropolitan area, studying both the willingness to pay for amenities and the inequality of the distribution of amenities over Paris. Although they do not estimate the influence of these amenities on house prices, the amenity variables used in this study are considered to have an influence on house prices (De Palma et al., 2007). De Palma et al. (2007) include variables related to accessibility to public transport, highways and arterial roads, employment and shops. Furthermore, public facilities, sports facilities, green areas and water are part of the analysis. Disamenities are considered as well, such as places that create noise pollution (Palma et al., 2007). Sedoarisoa et al. (2017) further specify noise pollution in traffic noise and aircraft noise. Baltagi et al. (2015) include the distance to the centre of the 'arrondissement' and to the centre of the quartier in their analysis, in which they find a significant negative relationship in one of the models.

Lastly, neighbourhood characteristics are characteristics that can be attributed to an entire neighbourhood, which are considered to have an effect on house prices as well. These characteristics determine the attractivity of the neighbourhood. Therefore, local amenities can be part of the neighbourhood, but there are other neighbourhood characteristics as well. The way in which neighbourhood characteristics are included in a model varies. The most straightforward way to correct for neighbourhood characteristics is by including location fixed effects, from coordinate to neighbourhood level, as is done by Mills et al. (2019), Cvijanovic & Spaenjers (2015), Cvijanovic & Spaenjers (2020), Gouriéroux & Laferrère (2009), Dubin (1992). Halvorsen & Pollakowski (1981) include the median income, median number of rooms, percent dwelling units owner-occupied and the employment accessibility index per census tract in San Francisco. Furthermore, Baltagi et al. (2015) mentioned the potential effect of crime rates and property taxes on house prices, but were not able to estimate the coefficients for these variables due to data limitations. Gravel et al. (2006) do include various property taxes, but also emphasize that these tax rates are only available on the city level. Furthermore, they include poverty rates in their analysis, defined as the fraction of households in the city exempt from property taxes (Gravel et al., 2006). De Palma et al. (2007) define redevelopment areas, which are characterised by social and economic difficulties. Lastly, in the studies about Paris, the street type is often defined. The categories are not identical, but avenue, boulevard, place and street are commonly used. Baltagi et al. (2015) found results with varying significance among different models, but all significant coefficients were positive. Therefore, a house located on an avenue, place or boulevard has a significantly higher house price than a house located on a street. Bresson & Hsiao (2011) find similar results and Amédée-Manesme et al. (2017) finds coefficients that vary in significance and size for these variables.

Table 1 House value drivers from theory France & Paris

House value drivers	Findings	Sources
Property characteristics		
Size	+	Allen et al., 2017; Maurer et al., 2004; Baltagi et al., 2015; Cvijanovic & Spaenjers, 2015; Gouriéroux & Laferrère, 2009; Dubin, 1992; Bresson & Hsiao, 2011; Dubé & Legros, 2014;
Number of bedrooms	+	Allen et al., 2017; Mills et al., 2019
Number of bathrooms	+	Allen et al., 2017; Mills et al., 2019; Baltagi et al., 2015; Violand & Simon, 2007; Gouriéroux & Laferrère, 2009; Dubin, 1992; Bresson & Hsiao, 2011; Dubé & Legros, 2014; Halvorsen & Pollakowski, 1981
Maid's room	+	Baltagi et al., 2015; Bresson & Hsiao, 2011
Garage, basement or attic	+	Maurer et al., 2014; Baltagi et al., 2015; Dubin, 1992; Dubé & Legros, 2014
Presence and type of outdoor space	+	Maurer et al., 2004; Baltagi et al., 2015; Cvijanovic & Spaenjers, 2015; Violand & Simon, 2007; Gouriéroux & Laferrère, 2009; Dubin, 1992; Bresson & Hsiao, 2011; Dubé & Legros, 2014; Gravel et al., 2006
Duplex/triplex apartment	+	Amédée-Manesme et al., 2017
Floor level	+/-	Maurer et al., 2004; Baltagi et al., 2015; Amédée-Manesme et al., 2017; Dubé & Legros, 2014
Presence of an elevator	+	Gouriéroux & Laferrère, 2009; Maurer et al., 2004; Cvijanovic & Spaenjers, 2015; Dubé & Legros, 2014
Parking places	+	Cvijanovic & Spaenjers, 2015; Dubin, 1992; Bresson & Hsiao, 2011
Construction year	+[1]	Maurer et al., 2004; Baltagi et al., 2015; Violand & Simon, 2007; Gouriéroux & Laferrère, 2009; Bresson & Hsiao, 2011; Dubé & Legros, 2014
Age of the construction	-	Allen et al., 2017; Mills et al., 2019; Halvorsen & Pollakowski, 1981;

[1] There are less recent construction periods that have a positive impact on property values. Therefore, the relationship is in most cases not linear.

Local amenities		
Natural amenities/green and water	+	Brueckner et al., 1997;
Historical amenities	+	Brueckner et al., 1997; De Palma et al., 2007
Proximity to public transport	+	De Palma et al., 2007
Proximity to highways	+	De Palma et al., 2007
Proximity to arterial roads	+	De Palma et al., 2007
Proximity to shops	+	De Palma et al., 2007
Distance to centre arrondissement, centre quartier & city centre	-	Baltagi et al., 2015
Noise pollution (disamenity)	-	De Palma et al., 2007; Sedoarisoa et al., 2017
Neighbourhood characteristics		
Location fixed effects (neighbourhoods, zip codes)	+/-	Mills et al., 2019; Cvijanovic & Spaenjers, 2015; Cvijanovic & Spaenjers, 2020; Gouriéroux & Laferrère, 2009; Dubin, 1992
Crime rates	NA	Baltagi et al., 2015
Property taxes	-	Baltagi et al., 2015; Gravel et al., 2006
Poverty rates	-	De Palma et al., 2007; Gravel et al., 2006
Street type	+/-	Baltagi et al., 2015; Bresson & Hsiao, 2011; Amédée-Manesme et al., 2017

2.2 Institutional investors and house prices

The relationship between investors and house prices has been studied by several authors. To begin with, (Allen et al, 2017; Smith & Liu, 2020; Bracke, 2013; Mills et al., 2019) study whether investors buy houses at a premium or discount compared to owner occupiers. Differently, Lambie-Hanson et al. (2019) and Mills et al. (2019) estimate the effect of residential investments on house price growth in the vicinity of the investment transaction. Furthermore, Allen et al. (2017) estimate house price levels instead of house price growth, with a similar key-independent variable. Several authors formulate a theory about the relationship between investor activity and house prices as well, based on other studies or qualitative research (Allen et al., 2017; Lambie-Hanson et al. 2015; Mills et al., 2019).

Several studies estimate differences in house prices between investor buyers and owner-occupiers and find that investors buy generally at a discount. All these studies estimate house prices by including one or several dummy variables for the buyer type, where the dummy indicates whether the buyer is an investor. Allen et al. (2017) find in their study of Miami between 2009 and 2013 that properties purchased by investors are purchased at a discount, compared to houses bought by owner-occupiers. They distinguish different investor categories based on their size, of which institutional investors buy relatively at the smallest discount. Smith & Liu (2020) conducted a similar study for Atlanta, United States with single-family detached housing transactions between 2000 and 2014. While institutional investors were found to buy at a discount of 6.3 to 11.8 percent after the real estate crisis, this discount was only 1.7 percent before the crisis (Smith & Liu, 2020). Conflicting with Allen et al. (2017), Smith & Liu (2020) find that institutional investors buy at a greater discount than non-institutional investors.

Mills et al. (2019) estimate house prices between 2000 and 2014 in the United States, by including dummies for investor types based on their size and a dummy for buy-to-rent investors. Investors classified as medium or large, are similar in size to investors classified as institutional investor in the paper of Allen et al. (2017). Results of the regression show that buy-to-rent investors purchase at a premium compared to individual investors, while corporate investors pay much less than individual investors (Mills et al., 2019). Bracke (2021) finds a divergent result for investors in general in his study about England and Wales in 2009 to 2014. The estimation of house prices with a key-independent dummy variable for buy-to-let purchases, shows a statistically significant discount of 1.6 and 3.9 percent for buy-to-let purchases (Bracke, 2021). However, the reference category is different as well, namely average prices. While the study period is equal to the time period of Allen et al. (2017), the housing market in England and Wales experienced different economic conditions, depending on the geographical area (Bracke, 2021). These studies imply that in general investors buy at a discount in comparison to owner occupiers and investors have a higher bargaining power. Their power in the market could result in investors influencing house prices in general, which is what we study in this paper.

Lambie-Hanson et al. (2019), Mills et al. (2019) and Allen et al. (2019) study the effect of residential investment purchases on house prices in a similar study area, in which they use the share of institutional investor purchases as key-independent variable. Lambie-Hanson et al. (2019) estimate house price growth rates on zip code-level between 2000 and 2014 in the United States, with net share of institutional investors as key-independent variable in the regressions. Furthermore, a 2SLS regression with an instrumental variable as replacement is conducted because of potential endogeneity. This suspicion for endogeneity originates from institutional investors responding to local economic conditions, resulting in growing investor activity when house prices are increasing. The first regression reveals a rise of real house prices of 1.2 basis points for every one-percentage-point increase of net share of institutional investors. The results of the 2SLS regression show a larger effect on house price growth rates. Furthermore, the effects of institutional investors on house price growth rates are larger in areas with small housing supply elasticity (Lambie-Hanson et al., 2019), which is the case in Paris as well (Roehner, 1999). Mills et al. (2019) use an identical dataset and dependent variable, but included share of buy-to-rent property transactions of total number of transactions per zip code as explanatory variable, instead of focusing on institutional investors. Similar to Lambie-Hanson et al. (2019), they use a linear regression and an instrumental variable approach, in order to deal with possible endogeneity because of investors entering a market where house prices are about to recover. Mills et al. (2019) find that investors contributed to higher house prices, with a significant effect in the years 2012 and 2014. More specifically, they find a 5-percentage point increase in buy-to-rent share leads to a 2-percentage point increase in house prices in this period. Mills et al. (2019) and Lambie-Hanson et al. (2019) both use house price growth as dependent variable, rather than house price levels. Therefore, their model looks different to our model, as they use control variables that vary over time instead of a hedonic model with property characteristics. Furthermore, both studies use an identical dataset and study the US recovering housing market. Therefore, they study whether investors contribute to the market recovery with their growing presence, instead of increasing the pressure in an already overheated housing market like Paris. Lastly, Allen et al. (2017) estimate house price levels instead of house price growth, with their study in Miami. In addition to their estimation including dummies for investor buyers, they estimate house price levels with the share of investor purchases. In their study, they argue that because investors are buying at a discount, they are not likely to be responsible for the housing market recovery. However, they do find a 10 percent-increase of the share of investor purchases resulting in a 0.20 percent increase in house prices in that census block (Allen et al., 2017). In their estimation, they use the share of investor purchases per census block as independent variable. Furthermore, they control for size and quality of a house and use location fixed effects on the census tract level. These results indicate that the presence of investors create upward pressure in the housing market. Allen et al. (2017) provide the explanation that the presence of properties for sale have negative externalities, which diminish when investors buy these properties. This argumentation is not applicable to the housing market of Paris, where there is a lack of supply instead of oversupply. Although we cannot compare these results to the findings of Lambie-Hanson et al. (2019) and Mills et al. (2019), Allen et al. (2017) find results in a similar direction. However, it should be stressed that all three studies use the relative number of investments or investors as regressor, instead of considering the size of the investments, which is what we do in this research. Furthermore, they use the share of investor purchases per administrative area, whereas we will calculate the investment values within radiuses around each transaction.

Finally, a selection of studies includes reasoning about the relationship between investor activity and house prices, based on theory and qualitative analyses. First of all, Allen et al (2017) mentions that in distressed markets, investors in general are presumably responsible for the overall price recovery, because of their high purchase power. Lambie-Hanson et al. (2015) mention their greater ability for financing homes as a reason for their active role in acquiring foreclosed properties in the Boston area. These actions resulted in absorbing a high volume of distressed properties and therefore investors had a stabilizing influence on this housing market (Lambie-Hanson et al., 2015). More specifically, Mills et al. (2019) argue that buy-to-rent properties are likely to have a larger effect on house prices, as these properties are removed from the buyer's market for a longer period of time, if not permanently. Therefore, investors buying properties to rent out, result in decreasing supply of houses for sale. Simultaneously, investors searching for houses to invest in, lead to increasing demand. Considering market fundamentals as described by Evans (2008) and the fact that supply cannot easily adjust in Paris due to space constraints, house prices are likely to rise as a result from increasing demand and decreasing supply. Mankiw & Weil (1989) study house prices in Paris between 1970 and 1980 and find supporting results. With house price as dependent variable, they find that a one percent increase in demand for housing, results in a 5.3 percent increase in the real price of housing.

2.3 Hypotheses

The first part of our theoretical framework consists of hedonic literature, mainly about house prices in Paris. The variables in the hedonic literature to estimate house prices can be distinguished into property characteristics, local amenities and neighbourhood characteristics and are summarized in table 1. With this study, we focus on the relationship between investor activity and house prices. Therefore, we will use the variables from hedonic studies to control for house price effects that can be attributed to the property characteristics instead of the investment activity.

Literature about investment activity and house prices is divided into studies estimating house prices with a dummy for the buyer type and studies estimating house prices or house price growth in the area with the share of investor purchases in that area. Investors are found to generally purchase houses at a discount. Although one could expect that with investors buying at discounts, average house prices would decrease with their presence, studies find that the share of purchases of investors has a positive relationship with house price growth and house price levels in the area. Therefore, we find substantial evidence that investor activity is associated with higher house prices, which is why we formulate the following hypothesis:

H1: There is a positive correlation between the share of institutional investments in housing and house prices.

The market for apartments and single-family homes can be considered different housing markets, as individuals often specifically search for either an apartment or a single-family home. Given the fact that the majority of the institutional investments are purchases of multi-family homes, investors operate mostly on the apartment market rather than the single-family home market. Relating to the supply and demand mechanisms, investor demand would cause the demand for apartments to rise more as well. Following this line of argumentation, investment activity is expected to have a larger effect on apartment prices than on prices of single-family homes. Accordingly, we formulate a second hypothesis:

H2: The correlation between institutional investments and apartment prices is stronger than the relationship between institutional investments and house prices.

Lastly, Lambie-Hanson et al. (2019) and Mills et al. (2019) state that the effect of investors is related to the elasticity of the housing supply. Considering that the inner city of Paris has a different population density and higher space constraints than the surrounding departments, it is likely that investors affect house prices in these areas differently. In order to assess whether this is the case, we formulate a third hypothesis about heterogeneity across the geographical departments.

H3: The correlation between institutional investments and house prices differs across departments.

3. Data & Method

3.1 Context

This study contains housing transactions within the greater Paris region, Île-de-France, with over 12 million inhabitants. Île-de-France is divided into 8 districts (départements), of which the district Paris is the inside the Paris ring road, containing 2.2 million inhabitants (Insee, 2021b). Paris is divided into 20 different neighbourhoods (arrondissements). The other districts of the greater Paris region are divided into communes, which are on the same aggregated level as the arrondissements (Britannica, 2021). The selection of districts is based on the investment activity in the departments. The small map in figure 2 displays the Greater Paris region and indicates the investor activity per district, determined by the number of transactions. The darker the grey colour, the higher the number of transactions. The district Paris and the first ring of districts situated directly around the inner city of Paris are included in the data analysis.

The study area of this research consists of four departments in the greater Paris region: Paris, Seine-Saint-Denis, Hauts-de-Seine and Val-de-Marne, which are displayed in figure 3. The average square meter prices are indicated as well, which are calculated from transactions in the sample.

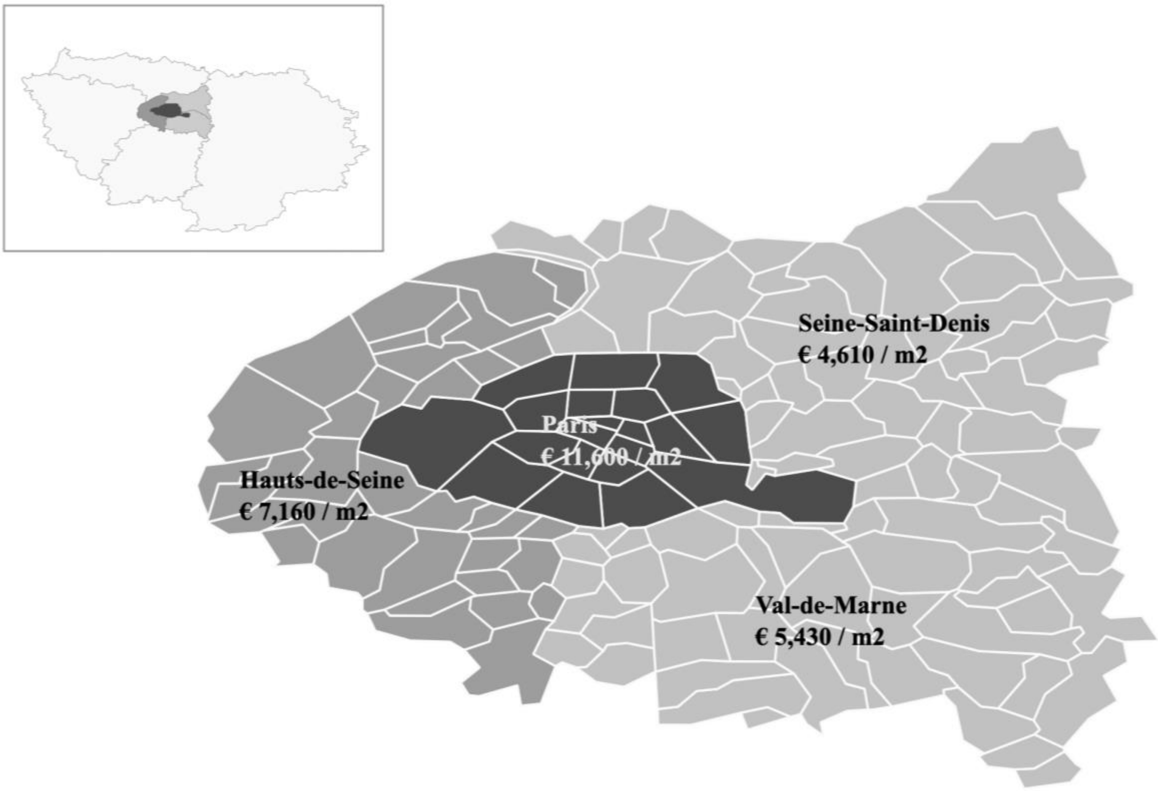

Figure 3 Map with study area and square meter prices (data from: Real Capital Analytics and Demande de Valeurs Foncières)[2]

[2] Due to confidentiality of the Real Capital Analytics investment data, the number of transactions are not displayed in the figure

3.2 Descriptive analysis

This study investigates the relationship between institutional investments in residential real estate and house prices in the areas of the investments, in the city of Paris. We combine publicly available French real estate transaction data from Demande de Valeurs Foncières (DVF), public transport data from RATP, data on public gardens and parks from both Mairie des Paris and Grand Paris Seine Ouest and data of residential real estate investments from Real Capital Analytics. The variables derived from these datasets are displayed in the descriptive statistics table in table 2 and a more elaborate description of the data is included in appendix I.

The dataset from DVF is modified and delivered by AEW, and contains 1,260,160 sale transactions of apartments or houses in the greater Paris region. From this dataset, we derive the transaction price, surface area, number of rooms, house type, location and the transaction date. First, we drop observations with missing values for the variables transaction value, surface area, street addresses and geographical coordinates, which are a total of 27,679 observations. Second, we exclude observations outside the district Paris and the first ring of districts around Paris, which leaves a dataset that includes Paris, Hauts-de-Seine, Seine-Saint-Denis and Val-de-Marne. Therefore, 559,020 observations from other departments are deleted, leaving 676,279 observations for the analysis. Third, in order to improve the distribution, observations with the lowest and highest 1 percent values of the transaction value, number of rooms, real size built and price per square meter. As a result, an additional 42,656 observations are removed from the dataset. Lastly, observations with identical transaction id's have an identical sale day and address. Moreover, many of the observations with identical transaction id's have an extremely high price and square meter price, indicating that the transaction value is for the complete building but the square meters for one apartment in this building. Additionally, part of these duplicate transaction ids are houses with gardens larger than 500 square meters, of which the gardens are separately administrated. Both the gardens and the houses have an identical sale price and cannot be distinguished in the sample. Due to the fact that it is not possible to only keep the correct values for these observations, we delete a total of 100,311 observations additionally. Therefore, 530,939 transactions are included in the house price estimations.

Second, we incorporate the distance to the closest public transport station in the analysis. We use location data from the Paris public transport organisation RATP, which includes bus stops, metro stations and train stations. However, the type of public transport is not specified, resulting in all types of public transport stops being treated equal in this study. We use both the coordinates from the RATP database and the house transactions to calculate the distance to the nearest public transport station for every transaction. The distance is measured in kilometres.

Third, a variable for the distance to the closest park or public garden is generated with data derived from Mairie des Paris and Grand Paris Seine Ouest. We generate coordinates from Shapefiles, using the centre points of the locations as coordinates and calculating the distance of the nearest park or public garden for both files separately. Subsequently, we keep only the lowest value of both calculations, which is the distance to the nearest park or garden in kilometres.

Lastly, for the key explanatory variable, we derive data on residential real estate transactions from the Real Capital Analytics database. These transactions are translated into total transaction values in the close vicinity of each transaction. In order to capture investments that are nearby but outside the administrative borders of the neighbourhood, we calculate the investment values within a radius of 1.5, 2 and 2.5 kilometres around each housing transaction. We chose the size of the radius based on the mean and median surface area of an arrondissement, where we determined what radius would approximately cover such an area. Furthermore, these radiuses were compared to other radiuses used for amenities in hedonic studies in Paris (Fack & Grenet; Bression & Hsiao, 2011). This variable contains the sum of the investment transactions between 2014 and 2020 within each radius. These values are absolute investment values in million euros over the complete period of 2014 to 2020 and therefore not dependent on the time of the housing transaction. The complexity of these calculations and the little amount of residential investment transactions made a time depending radius calculation an unavailable option for this study.

Table 2. Descriptive statistics

Variable	Mean	Std. Dev.	Min	Max
transaction value	437928.32	1193967.8	15000	1.655e+08
investment value radius 1500 m	110.427	131.038	0	728.818
investment value radius 2000 m	197.042	203.652	0	957.208
investment value radius 2500 m	305.728	293.106	0	1250.046
investment value radius 1500-2000 m	86.615	108.739	0	666.364
investment value radius 2000-2500 m	108.686	128.808	0	869.135
surface area of the house (m^2)	58.586	31.402	10	181
number rooms	2.732	1.272	1	7
single-family home (dummy, yes=1)	.136	.343	0	1
sale year	2017.058	1.944	2014	2020
distance closest public transport stop (km)	.177	.39	.001	8.045
distance closest park (km)	1.753	2.478	0	14.819
Number of observations 530,939				

The descriptive statistics of the variables included in the analysis are listed in table 2. The transaction value is included as continuous variable. It should be noted that the mean transaction value is likely to be slightly higher in reality, due to the fact that all houses with gardens larger than 500 square meters are deleted from the dataset. Nevertheless, the average square meter prices per district, as indicated by figure 3, are similar to the average square meter prices from different sources (Chambre des Notaires de Paris, 2021; SeLoger, 2021). In addition to the houses with a garden larger than 500 square meters, the observations with the highest and lowest 1 percent transaction value were deleted as well. However, the range of the transaction value is still large, which suggests that the housing market in Paris is heterogeneous. The investment values are included as continuous variables as well and is measured with three radiuses. Logically, the larger the radius, the higher the average investment value, as more investment transactions are included in the radius. The standard deviations are high, indicating that the investment activity in the close vicinity varies relatively much for different housing transactions. The surface area of the house is again included as continuous variable. For the surface area, we delete the observations with the highest and lowest 1 percent values as well. We delete the highest and lowest 1% of the values for number of rooms additionally, but we include the number of rooms as a dummy variable in the analysis. Furthermore, we include a dummy variable for the house type, where we distinguish apartments and single-family homes. In order to control for variations of prices over time, we include the sale year, again separated into dummy categories. For each house transaction, we include the distance to the closest public transport stop and the distance to the closest park as continuous variables. Every location from these datasets is included in the distance calculations, but when certain locations are not registered, it can result in an incorrect distance measurement.

Table 3. Correlation matrix

	Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	transaction value	1.000									
(2)	surface area (m²)	0.176	1.000								
(3)	distance closest public transport stop (km)	-0.024	0.130	1.000							
(4)	distance closest park (km)	-0.078	0.216	0.496	1.000						
(5)	number of rooms	0.128	0.860	0.139	0.260	1.000					
(6)	investment value radius 1500 m	0.083	-0.144	-0.157	-0.456	-0.190	1.000				
(7)	investment value radius 2000 m	0.088	-0.168	-0.179	-0.519	-0.215	0.877	1.000			
(8)	investment value radius 2500 m	0.090	-0.186	-0.192	-0.555	-0.231	0.805	0.928	1.000		
(9)	investment value radius 1500-2000 m	0.064	-0.142	-0.146	-0.423	-0.173	0.438	0.816	0.768	1.000	
(10)	investment value radius 2000-2500 m	0.065	-0.157	-0.153	-0.442	-0.187	0.445	0.531	0.808	0.458	1.000

Table 3 indicates the correlation between the numerical variables included in the analysis. First of all, there is a high correlation between the different investment variables, which can be explained by the fact that these have overlapping data. In order to solve the overlapping data issue for these variables, two new variables are generated that contain only the differences between the investment values of the different ranges. Logically, these new generated variables have a substantially lower correlation with each other and with the variable that indicates the investment value within a radius of 1500 m. We find a strong correlation between the variables number of rooms and the surface area as well, which is expected due to the fact that both represent the size of the house. These results indicate a potential multicollinearity issue, but we do not find any multicollinearity problems within the regression. Based on the correlation table, we do not expect other multicollinearity issues either. Additionally, we find a negative correlation between the investment variables and the variables that represent the distance to the closest public transport stop and distance to the closest park. This suggests that more is invested in areas where houses are close to these amenities. However, these relationships are not strong and studying this effect is beyond the scope of this research.

3.3 Models

For this study, we use the hedonic model from Rosen (1974) to estimate house price levels. This baseline model estimates house prices with a vector of house value drivers, divided into housing characteristics, local amenities and neighbourhood characteristics. Accordingly, the model based on theory is as follows:

$$P = f(H, L, N)$$

P stands for house price, which is determined by housing characteristics (H), local amenities (L) and neighbourhood characteristics (N). In our models, housing characteristics are defined by the surface area, number of rooms and house type. Furthermore, we include the distance to the closest public transport stop and distance to the closest park or public garden to account for the proximity to local amenities. Furthermore, we control for neighbourhood characteristics with location fixed effects on the commune level. Considering that house prices vary over time, we control for time with the sale year. With these variables, we estimate our baseline model:

$$\begin{aligned} \log(P) = {} & \beta_0 + \beta_1 \log(Surface\ area)_i + \Sigma_{r=2}^{7} \beta_2\ Number\ of\ rooms_i \\ & + \beta_3 dSingle\ family\ home_i + \beta_4 \log(Distance\ Public\ Transport)_i \\ & + \beta_5 \log(Distance\ Park)_i + \Sigma_{n=2}^{143} \beta_6 Neighbourhood_{i,n} \\ & + \Sigma_{t=2}^{6} \beta_7 Sale\ year_{i,t} + \varepsilon_{i,t} \end{aligned} \quad (1)$$

We estimate the natural logarithm of the transaction value ($\log(P)$) to better approach a normal distribution and mitigate the effect of outliers of the dependent variable. For the surface area we include $\log(Surface\ area)_i$, which is the natural logarithm of the surface area in square meters. $Number\ of\ rooms_i$ is a vector of seven dummies, where the value of 1 indicates that observation i has r number of rooms and the value of 0 otherwise. The variable $dSingle\ family\ home_i$ is a dummy that takes on the value of 1 if the house type is a house and of 0 if it is an apartment. $\log(Distance\ Public\ Transport)_i$ takes the natural logarithm of the distance in kilometres to the nearest public transport station, which can be a bus, metro or train station. Subsequently, $\log(Distance\ Park)_i$ is the natural logarithm of the distance in kilometres to the nearest park or public garden. For the location fixed effects, a vector of 143 dummies is included where observation i takes the value of 1 when the house is located in neighbourhood n and 0 otherwise. Lastly, time fixed effects are included with a vector of six dummies for the sale year, where i takes the value of 1 when the transaction took place in year t and 0 otherwise.

The key explanatory variable is $Investment\ value_{i,r}$, which is the sum of the absolute investment values of each investment transaction between 2014 and 2020 in million euros. These investment values are measured within a radius of 1500, 2000 and 2500 metres around each housing transaction. We did not transform the investment values into natural logarithms, because in a substantial number of observations, the investment variable takes the value of zero. Therefore, the following model is estimated three times, of which one time for each radius.

$$\begin{aligned} \log(P) = \beta_0 + \beta_1 Investment\ value_{i,r} + \beta_2 \log(Surface\ area)_i \\ + \Sigma_{r=2}^{7} \beta_3\ Number\ of\ rooms_i + \beta_4 dSingle\ family\ home_i \\ + \beta_5 \log(Distance\ Public\ Transport)_i + \beta_6 \log(Distance\ Park)_i \\ + \Sigma_{n=2}^{143} \beta_7 Neighbourhood_{i,n} + \Sigma_{t=2}^{6} \beta_8 Sale\ year_{i,t} + \varepsilon_{i,t} \end{aligned} \tag{2}$$

The most complete model includes total investment value within a radius of 1500 meters, total investment value in between the radiuses of 1500 and 2000 meters and total investment value in between a radius of 2000 to 2500 meters.

$$\begin{aligned} \log(P) = \beta_0 + \beta_1 Investment\ value_{i,r=1500} + \beta_2 Investment\ value_{i,r=2000-1500} \\ + \beta_3 Investment\ value_{i,r=2500-2000} + \beta_4 \log(Surface\ area)_i \\ + \Sigma_{r=2}^{7} \beta_5\ Number\ of\ rooms_i + \beta_6 dSingle\ family\ home_i \\ + \beta_7 \log(Distance\ Public\ Transport)_i + \beta_8 \log(Distance\ Park)_i \\ + + \Sigma_{n=2}^{143} \beta_9 Neighbourhood_{i,n} + \Sigma_{t=2}^{6} \beta_{10} Sale\ year_{i,t} + \varepsilon_{i,t} \end{aligned} \tag{3}$$

The models are estimated using an Ordinary Least Squares (OLS). In order to make statistical inferences based on the results, five assumptions should be met (Brooks & Tsolacos, 2010). These assumptions are discussed in detail in appendix II.

Furthermore, we test the extent to which our estimated coefficients for the investments are subjective to a change in the location fixed effects. It is possible that the neighbourhood dummies absorb part of the effect of investments, as the investment value in the close vicinity is related to the location of the house. Therefore, we assess the robustness of the models by estimating two additional models with location fixed effects on both a lower and higher aggregated level. The fourth model includes the location fixed effects with dummies for the streets and the fifth model includes a vector of four dummies for the districts.

4. Results

In order to gain insight in the correlation between institutional investments and house prices in the close vicinity, we first estimate the five models in table 4. Model 1 is our baseline hedonic model, without the inclusion of the key-independent variable. Model 2, 3 and 4 all include one of the radius variables separately. Lastly, model 5 is our most complete model, where we include the smallest radius and the additional investment values by including investments within the two larger radiuses as well.

Table 4. Regressions estimating log(P)

	(1)	(2)	(3)	(4)	(5)
Surface area of the house	.84459***	.844***	.84392***	.8439***	.84366***
	(.00246)	(.00246)	(.00246)	(.00246)	(.00246)
Single-family home (yes=1)	.30995***	.31083***	.3112***	.31121***	.31157***
	(.00209)	(.00209)	(.00209)	(.00209)	(.00209)
1 bedroom (reference)					
2 bedrooms	.01678***	.01744***	.01738***	.01743***	.01769***
	(.00219)	(.00219)	(.00219)	(.00219)	(.00219)
3 bedrooms	.03893***	.0399***	.03987***	.04003***	.04039***
	(.00273)	(.00273)	(.00273)	(.00273)	(.00273)
4 bedrooms	.03603***	.03693***	.03681***	.03702***	.03734***
	(.00338)	(.00338)	(.00338)	(.00338)	(.00338)
5 bedrooms	.03612***	.03668***	.03629***	.03647***	.03666***
	(.00415)	(.00415)	(.00415)	(.00415)	(.00415)
6 bedrooms	.01935***	.01966***	.01941***	.01949***	.0196***
	(.00542)	(.00542)	(.00541)	(.00541)	(.00541)
7 bedrooms	.00614	.00689	.00699	.00694	.00726
	(.00851)	(.0085)	(.0085)	(.0085)	(.00849)
Distance closest public transport stop	-.00114	-.00145	-.0016*	-.00191**	-.00196**
	(.00089)	(.00089)	(.00089)	(.00089)	(.00089)
Distance closest park	-.0111***	-.01056***	-.01027***	-.00944***	-.00943***
	(.00097)	(.00097)	(.00097)	(.00097)	(.00097)
Investment value radius 1500 m		.00017***			.00023***
		(.00001)			(.00001)
Investment value radius 2000 m			.00017***		
			(.00001)		
Investment value radius 2500 m				.00015***	
				(.00001)	
Investment value radius 1500-2000 m					.00015***
					(.00001)
Investment value radius 2000-2500 m					.00011***
					(.00001)
Observations	530939	530939	530939	530939	530939
R-squared	.69705	.69733	.69749	.69762	.6977
Location fixed effects	YES	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES	YES

Dependent variable is the natural log of transaction prices
Location fixed effects are on neighbourhood level
Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

The first model does not include our key-independent variables, but serves as a reference for the other models. Therefore, we will not interpret the separate coefficients of this model. For the other four models, we find positive coefficients for all investment variables, which are all significant on the 1 percent level. Therefore, these results suggest that the value of residential investments in the close vicinity indeed have a positive association with house prices. Furthermore, we find that the larger the investment radius around the house transaction, the lower the coefficient is. Additionally, we find that the additional investments within a radius of 2000 to 2500 meters have a lower coefficient than the additional investments in the radius between 1500 and 2000 meter. These coefficients suggest that the strength of the correlation between investment values and house prices decreases with the distance of the investments from the house. Although all coefficients are significant, they are rather small. We find that every 1 million euro increase of the investment value within the radiuses is associated with at most a 0.0225% higher house price.[3] Therefore, the question can be asked whether these results are actually economically significant. In order to provide insight in this matter, we calculate the average house price increases. The average investment values for the three radiuses are approximately 110, 197 and 305 million euros. For the coefficients of model 2, 3 and 4, this calculation results in house prices that are approximately 1.87 percent, 3.27 percent and 4.6 percent higher. For our last model, we make the house price calculation for each radius again. Using the same average investment values, we still have an investment value of 110 million euros within a radius of 1500 meters. For the extra 500-meter radius we have an added investment value of 87 million euros and another 110 million euros added for investments outside a radius of 2000 meters but within 2500 meters. By filling in these values in model 5, we get 2.47, 1.35 and 1.71 percent higher house prices. The results from this final model indicate that house prices are on average 5.53 percent higher in areas with average investment activity. These results imply substantially higher house prices for houses with residential investments in the close vicinity compared to houses that do not have investments in the close vicinity.

4.1 Robustness

We consider the possibility that the location fixed effects absorb part of the investment coefficients, as the investment value within the close vicinity is dependent on the location of the house transaction. Because the investment variable is included as a time-invariant variable, it only contains the locations of the investments and the amount invested. Therefore, these investments become location fixed effects as well, possibly resulting in a higher degree of overlap with the neighbourhood fixed effects. In order to explore this issue, we assess the robustness of the coefficients by varying the level of the location

[3] The increases are calculated for every investment coefficient with the same approach of Halvorsen and Palmquist (1980). This results in the percentage values 0.0170; 0.0167; 0.0151; 0.0225; 0.0155; 0.0114.

fixed effects from neighbourhood into districts and streets. The results in table 5 show the coefficients for the models with the three types of location fixed effects.

Table 5. Summary of estimating log(P) with different aggregated levels

	(1) Neighbourhoods	(2) Districts	(3) Streets
Investment value radius 1500 m	.00023***	.00035***	.0007***
	(.00001)	(.00001)	(.00001)
Investment value radius 1500-2000 m	.00015***	.00023***	.00047***
	(.00001)	(.00001)	(.00001)
Investment value radius 2000-2500 m	.00011***	.00022***	.00028***
	(.00001)	(.00001)	(.00001)
Observations	530939	530939	530939
R-squared	.6977	.63425	.6651
Property characteristics	YES	YES	YES
Location fixed effects	YES	YES	YES
Time fixed effects	YES	YES	YES

Dependent variable is the natural log of transaction prices
Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

The coefficients of the key-independent variables all remain positive and significant at the 1 percent-level. However, the coefficients are larger in both the model with location fixed effects on the district level and on the street level. This suggests that the neighbourhood fixed effects absorb part of the coefficients of the investments. However, it should be noted that with these different location fixed effects, the investment coefficients potentially absorb part of the neighbourhood fixed effects as well. Therefore, it can be expected that in reality the correlation between house prices and investment values in the close vicinity is in between these values.

4.2 Heterogeneity

Investors buy mostly multi-family apartment buildings. Therefore, we can expect that their demand for apartments is stronger related to apartment prices than to prices of single-family homes. In order to test this suspicion, we estimate two separate models for apartments and houses. The results are given in table 6. Furthermore, summary statistics are included in appendix III.

Table 6. Regressions estimating log(P) for apartments and houses

	(1) Single-family homes	(2) Apartments
Investment value radius 1500 m	.00018***	.00022***
	(.00007)	(.00001)
Investment value radius 1500-2000 m	.00027***	.00014***
	(.00006)	(.00001)
Investment value radius 2000-2500 m	.00009*	.00011***
	(.00005)	(.00001)

Observations	72290	458649
R-squared	.64921	.70462
Property characteristics	YES	YES
Location fixed effects	YES	YES
Time fixed effects	YES	YES

Dependent variable is the natural log of transaction prices
Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

In the separate regressions for apartments and houses, the investment variables still each have significant and positive coefficients in both estimations. Therefore, we find that for both house types there is a positive correlation with investment activity. However, the coefficients differ in value, implying that the strength of the correlation between investments and house prices differs for apartments and single-family homes. In order to test whether the coefficients of the investment variables are actually significantly different, we perform a Chow-test. The Chow test result indicates an F-score of 1,122.57[4] which is higher than the critical value for 1 percent significance. Based on these results, we conclude that the correlation between investment value and house prices is significantly different for apartments and houses. Therefore, the pooled model to estimate house prices for both apartments and single-family homes is not sufficient.

Second, to determine whether we can generalise our results and can use our pooled model for the whole study area, we test for heterogeneity across the four geographical departments. The results of the four separate regressions are displayed in table 7 and summary statistics are included in appendix III.

Table 7. Regressions estimating log(P) for the geographical departments

	(1) Paris	(2) Hauts-de-Seine	(3) Seine-Saint-Denis	(4) Val-de-Marne
Investment value radius 1500 m	.00022***	-.00002	.00045***	.00004
	(.00001)	(.00002)	(.00004)	(.00005)
Investment value radius 1500-2000 m	.00014***	.00019***	.00015***	-.00009***
	(.00001)	(.00002)	(.00003)	(.00003)
Investment value radius 2000-2500 m	.00013***	.00001	.00012***	.00003
	(.00001)	(.00002)	(.00003)	(.00003)
Observations	207445	127406	90496	105592
R-squared	.69455	.69915	.52649	.64632
Property characteristics	YES	YES	YES	YES
Location fixed effects	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES

Dependent variable is the natural log of transaction prices
Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

[4] F = ((80,934.53-71,017.8325-8,240.8575)/(71,017.8325+8,240.8575))*((530,939-2*10)/(2*10-10)) = 1,122.57

The regression output indicates insignificant coefficients in the districts Hauts-de-Seine and Val-de-Marne for the variables investment value within 1500 metres and investment value between 2000 and 2500 metres. These districts are not specifically the departments with the lowest square meter prices or the lowest investor activity according to figures 2 and 3. For the departments Paris and Seine-Saint-Denis, we still find that all coefficients are significant at the 1 percent-level and higher coefficients for the investments nearby. In order to assess whether the coefficients are heterogeneous across the districts, we perform a Chow-test here as well. We calculate an F-statistic of 494.95[5], which is higher than the critical value for an F-distribution with a 99% confidence level. Therefore, we conclude that the coefficients for the investment values differ significantly across the four geographical districts in the sample.[4] To conclude, we improve the fit of the model significantly by estimating four separate regressions for the different districts.

5. Discussion

The main purpose of this research is to gain insight in the relationship between institutional investments in residential real estate and house prices in the close vicinity. We assess whether there is a positive correlation between these investments and house prices and if these results are generalisable for the study area and different house types.

The results suggest that there is indeed a positive correlation between house prices and the value of institutional investments in the close vicinity. We find very significant positive coefficients for every key-independent variable in our first set of models. Additionally, we find lower coefficients for the investment variables with larger radiuses, implying that house prices have a less strong correlation with investments further away. Furthermore, we find that the coefficients change when the aggregated level of the location fixed effect changes. We expect that part of the coefficients that can be attributed to the investments is absorbed by the location fixed effects, with the implication that the correlation between house prices and investment values is expected to be stronger in reality. Lastly, the evidence we examine suggests that there is heterogeneity across house types and districts in Paris. Therefore, we conclude that the correlation between house prices and investment values is not generalisable for different house types and different areas.

The central question about the correlation between house prices and institutional investments has been answered with the study, as results report significant and positive correlations between investment values and house prices. The data supports all three hypotheses. First of all, we find a positive correlation between the investments and house prices. Second, the results provide evidence that there is a

[5] F = ((80,934.53-35,318.3581-16,483.0349-14,071.005-12,860.0969)/(35,318.3581+16,483.0349+14,071.005+12,860.0969))*((530,939-4*10)/(4*10-10)) = 494.95

significantly stronger correlation between investments and apartment prices than between investments and prices of single-family homes. Lastly, coefficients differ significantly between the included districts in the Greater Paris region. However, although we find the expected heterogeneity across departments, the division of the coefficients across the departments is unexpected. We would expect that the correlation relates to the circumstances of the specific housing market. However, the results do not suggest that there is a linear relationship between the coefficients of the investment variables and the average square meter prices of the different departments.

The empirical evidence is consistent with findings of Lambie-Hanson et al. (2019), Mills et al. (2019) and Allen et al. (2017), which all find a positive significant relationship between investments and house prices or house price growth. However, it should be noted that it is not possible to directly compare these results as there is a difference in modelling and study area. First of all, Lambie-Hanson et al. (2019) and Mills et al. (2019) estimate house price growth instead of house price levels. Second, all three studies include investor activity as a variable that indicates the share of number of investments relative to total number of transactions. This variable includes number of investments, instead of the total investment values we use in this study. Additionally, these studies estimate the effect of investments on house prices, while this study only indicates a correlation instead of causation. Furthermore, we use investment value within a radius around each transaction, instead of the share of investments in a certain administrative area. Considering that we find heterogeneity across different departments in Paris already, we should consider as well that all studies are conducted in the United States, with different economic circumstances.

This study contributes to the literature with its approach of measuring investment activity. Because we use radius calculations to measure investment activity, we look beyond administrative borders and use an equally sized searching area for each transaction. Furthermore, we consider the magnitude of the investments by including the value of the investment transactions. Institutional investors often make large transactions, buying multiple apartment units in one transaction. By including total investment value instead of number of transactions, we distinguish transactions of different sizes. However, it should be noted that a combination of both the approach of this study and previous studies is preferred, considering the share of investment value relative to the total transaction value. Second, we add to the knowledge in a new study area, whereas previous research on this topic is heavily focused on the United States. Our findings suggest that there is a significantly different correlation in different areas in Paris and previous studies find differences depending on market circumstances (Lambie-Hanson et al., 2019; Mills et al., 2019; Allen et al., 2017). This emphasizes the importance of doing local research and of not making inferences based on previous studies with different study areas and market circumstances.

Nevertheless, there are limitations regarding the data and model in this research that should be considered. The most important limitation is related to the investment variable, which we did not include in the envisioned manner. In order to estimate a causal effect where house prices are higher due to higher investment activity, the investment value variable should only include investments made in a certain period before the house transaction. However, due to data and time limitations, we were only able to include a variable that measures the investment value of the complete period of 2014 to 2020. Therefore, we include investments in house price estimations that could not have impacted specific transactions, simply due to the fact that the investment transaction happened in the years after the house transaction. The time-independent investment variables that we included in our analysis can be interpreted as a measurement of an overall investment activity level. Because we are not able to determine whether house prices or investment levels rose first, we can only refer to our results as correlations between our dependent and explanatory variable. Furthermore, we include the absolute investment values in the model instead of a share of investment value out of total transaction value. Therefore, we do not consider the total activity of all buyers on the Paris housing market. Including the investment values as a share of total transaction value would give a more nuanced result, similar to the number of investments variables from Lambie-Hanson et al. (2019), Mills et al. (2019) and Allen et al. (2017). Second, the sample selection can be considered biased. The DVF database with housing transactions distinguishes transactions with individual transaction ids. However, gardens larger than 500 square meters are reported separately, but with an identical transaction id and transaction price to the house. Because we could not merge the transactions of the house and the garden or filter the gardens out of the dataset, we had to remove all identical transaction id's and exclude houses with a garden larger than 500 square meters from the sample. Furthermore, the housing transaction database of DVF does not provide construction years and the presence of outdoor space in their data. Considering that these variables have significant results in many hedonic studies, there is possible omitted variable bias. Additionally, our variables related to local amenities could be more nuanced. We do not consider differences between different types of public transport, whereas it is likely that these effects do differ. Similarly, we do not consider the size of the parks, which could have implications for the effect on house prices as well.

6. Conclusion

With this study, we aim to attribute to the knowledge available on the relationship between investors and house prices. We estimate house prices with a hedonic model and use an OLS regression for our estimation. Our findings provide evidence that the value of residential investments is positively correlated with house prices in the close vicinity.

The knowledge obtained from this paper can be useful in the decision-making process for housing policies and government intervention measures with regards to investors. Furthermore, this information is relevant for the investment strategies of institutional investors as well, as the information can be used to adjust their location strategies on. Lastly, the study has important scientific implications as well, whereas this study can form a basis for future research on this topic and emphasizes the importance of market specific research.

It is evident that this subject deserves further empirical study. Additional research is necessary to further specify the investment variable, by incorporating relative investment values that are time dependent. Another promising line of research would be to use a repeat sales model to estimate the effect, to control for omitted variable bias. Furthermore, it is interesting to further elaborate on the baseline hedonic model, by including more relevant variables. Additionally, it is expected that by distinguishing certain variables into separate categories, such as public transport and park distance in this case, the model would gain explanatory power. Lastly, repeating this research for different geographical areas can help to gain insight in the market conditions that relate to the difference of the effect of investments on house prices. This study forms a basis for the future research on this subject and serves as a precursor to more extensive research, in order to increase the knowledge on this topic.

References

AEW. (2020). Home Sweet Home – The Residential Sector Lies on Solid Foundations. *AEW*. Reviewed 27 February 2021, from: https://www.aew.com/research/european-flash-report-covid-19-special-update-6

Allen, M. T., Rutherford, J., Rutherford, R., & Yavas, A. (2017). Impact of investors in distressed housing markets. *The Journal of Real Estate Finance and Economics*, *56*(4), 622-652.

Amédée-Manesme, C. O., Baroni, M., Barthélémy, F., & Des Rosiers, F. (2017). Market heterogeneity and the determinants of Paris apartment prices: A quantile regression approach. *Urban Studies*, *54*(14), 3260-3280

Baléo, M. (2019). Paris: can densification rescue affordable housing? *La Fabrique de la Cité*. Retrieved from: https://www.lafabriquedelacite.com/en/publications/paris-can-densification-rescue-affordable-housing/

Baltagi, B. H., Bresson, G., & Etienne, J. M. (2015). Hedonic housing prices in Paris: an unbalanced spatial lag pseudo-panel model with nested random effects. *Journal of Applied Econometrics*, *30*(3), 509-528.

Bourla. (2021). How is the Paris residential market performing? *Knight Frank*. Reviewed 23 March 2021, from: https://www.knightfrank.com/research/article/2021-03-08-how-is-the-paris-residential-market-performing

Bracke, P. (2021). How much do investors pay for houses? *Real Estate Economics*, *49*(S1), 41-73.

Bresson, G., & Hsiao, C. (2011). A functional connectivity approach for modelling cross-sectional dependence with an application to the estimation of hedonic housing prices in Paris. *Advances in Statistical Analysis*, *95*(4), 501-529.

Britannica. (2021). Paris [online]. Retrieved from: https://www.britannica.com/place/Paris

Brooks, C., & Tsolacos, S. (2010). *Real estate modelling and forecasting*. Cambridge University Press.

Brueckner, J. K., Thisse, J. F., & Zenou, Y. (1999). Why is central Paris rich and downtown Detroit poor? An amenity-based theory. *European Economic Review*, *43*(1), 91-107.

Chambre des Notaires de Paris. (2021). Carte des prix de l'immobilier du Grand Paris | Fin Mars 2021. Retrieved from: https://paris.notaires.fr/fr/carte-des-prix

Cvijanovic, D., & Spaenjers, C. (2015). Real estate as a luxury good: Non-resident demand and property prices in Paris. HEC Paris Research Paper No. FiN-2015-1073.

Cvijanović, D., & Spaenjers, C. (2020). "We'll Always Have Paris": Out-of-Country Buyers in the Housing Market. *Management Science*. Published online.

Daams, M. N., Sijtsma, F. J., & van der Vlist, A. J. (2016). The effect of natural space on nearby property prices: accounting for perceived attractiveness. *Land Economics*, *92*(3), 389-410.

De Palma, A., Motamedi, K., Picard, N., & Waddell, P. (2007). Accessibility and environmental quality: inequality in the Paris housing market. *European Transport, 36*, 47-74.

Delmendo, L.C. (2020). France's housing market remains resilient. Global Property Guide. Reviewed 27 February 2021, from: https://www.globalpropertyguide.com/Europe/France/Price-History

Dubé, J., & Legros, D. (2014). Spatial econometrics and the hedonic pricing model: what about the temporal dimension*? Journal of Property Research*, 31(4), 333-359.

Dubin, R. A. (1992). Spatial autocorrelation and neighbourhood quality. *Regional Science and Urban Economics*, 22(3), 433-452.

Evans, A. W. (2008). *Economics, real estate and the supply of land*. John Wiley & Sons.

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.

Gouriéroux, C., & Laferrère, A. (2009). Managing hedonic housing price indexes: The French experience. *Journal of Housing Economics*, *18*(3), 206-213.

Gravel, N., Michelangeli, A., & Trannoy, A. (2006). Measuring the social value of local public goods: an empirical analysis within Paris metropolitan area. *Applied Economics*, *38*(16), 1945-1961.

Halvorsen, R., & Palmquist, R. (1980). The interpretation of dummy variables in semilogarithmic equations. *American Economic Review*, *70*(3), 474-475.

Halvorsen, R., & Pollakowski, H. O. (1981). Choice of functional form for hedonic price equations. *Journal of Urban Economics, 10*(1), 37-49.

Indiana University. (2018). In Stata, how do I conduct the Chow Test? [online]. Retrieved from: https://kb.iu.edu/d/chow

Insee. (2021a). Price index of second-hand dwellings – Paris – Flats – Base 100 = annual average of year 2015 – SA series. Insee. Reviewed 27 February 2021, from: https://www.insee.fr/en/statistiques/serie/010567013

Insee (2021b). Dossier complet Région d'Île-de-France (11) [online]. Retrieved from: https://www.insee.fr/fr/statistiques/2011101?geo=REG-11

Lambie-Hanson, L., Herbert, C. E., Lew, I., & Sanchez-Moyano, R. (2015). Foreclosed property investors in a strong housing market city: A case study of Boston. *Cityscape*, *17*(2), 239-268

Lambie-Hanson, L., Li, W., & Slonkosky, M. (2019). Leaving Households Behind: Institutional Investors and the US Housing Recovery. *Federal Reserve Bank of Philadelphia, 19*(1).

Laurent, L. (2020). Forget London House Prices. Take a Look at Paris. Bloomberg. Reviewed 28 February 2021, from: https://www.bloomberg.com/opinion/articles/2020-02-17/europe-s-superstar-cities-are-choking-local-home-buyers

Mankiw, N. G., & Weil, D. N. (1989). The baby boom, the baby bust, and the housing market. *Regional Science and Urban Economics*, *19*(2), 235-258.

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

Mills, J., Molloy, R., & Zarutskie, R. (2019). Large-scale buy-to-rent investors in the single-family housing market: The emergence of a new asset class. *Real Estate Economics*, *47*(2), 399-430.
Palmquist, R. B. (2005). Property value models. *Handbook of environmental economics*, *2*, 763-819.

Ministère de la Transition Écologique. (2021). House Prices in France: Property Price Index, French Real Estate Market Trends in the Long Run [online]. Retrieved from: http://www.cgedd.developpement-durable.gouv.fr/house-prices-in-france-property-price-index-french-a1117.html

Roehner, B. M. (1999). Spatial analysis of real estate price bubbles: Paris, 1984–1993. *Regional Science and Urban Economics*, *29*(1), 73-88

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 35-55.

Savills. (2020). Report: Paris Prime Residential Market 2020. Retrieved from: https://www.savills.com/research_articles/255800/295846-0

Sedoarisoa, N., Desponds, D., Dalmas, L., & Lavandier, C. (2017). Housing prices and the proximity of the Paris–Charles-de-Gaulle (CDG) airport. L*'Espace geographique*, *46*(1), 61-78.

SeLoger. (2021). Prix de vente au m^2 [online]. Retrieved from: https://www.seloger.com/prix-de-l-immo/vente/ile-de-france/paris.htm

Smith, P. S., & Liu, C. H. (2020). Institutional Investment, Asset Illiquidity and Post-Crash Housing Market Dynamics. *Real Estate Economics*, *48*(3), 673-709.

Statista. (2021). Average price per square meter of an apartment in Europe H1 2020, by city. Statista. Reviewed 27 February 2021, from: https://www.statista.com/statistics/1052000/cost-of-apartments-in-europe-by-city/

Strangler, C. (2020). The Death of Working-Class Paris. The Nation. Reviewed 28 February 2021, from: https://www.thenation.com/article/world/paris-gentrification/

The Economist. (2020). Which cities have the highest cost of living? The Economist. Reviewed 27 February 2021, from: https://www.eiu.com/n/campaigns/worldwide-cost-of-living-2020/

Violand, W., & Simon, A. (2007). Real estate brokers: Do they inflate housing prices. The case of France. *Concurrence & Consommation*, 154.

Appendix I. Data sources empirical research

For this study, we combine several databases. This appendix contains an elaborate description of the data sources used in this thesis.

Housing transactions

We use housing transactions from the publicly available data of Demande de Valeurs Foncières, which is a national database for property transactions and has been made available by the French government. For this study, we include transactions from the period 2014 to 2020 in the Greater Paris region.

Investment transactions

We derive investment values from the real estate investment database from Real Capital Analytics, which is not publicly available. Therefore, this data is transformed into the investment values by the global investment manager AEW. Consistent with the housing transactions, we use investment transactions from the years 2014 to 2020. The dataset contains residential investment transactions of purchases made by institutional investors.

Public transport

For the location of public transport stops, we use the most recent dataset from the local transport organisation RATP, which contains locations of metro stations and bus stops in the Greater Paris region. This dataset contains only the names and coordinates of each station, without making a distinction between bus stops and metro stations. This data is provided on the website of the organisation www.RATP.fr.

Parks and public gardens

For the locations of parks and public gardens, we combine two data sources. Instead of using the datasets with coordinates, we calculate new coordinates from the also available Shapefiles and use the centre points of each green space to derive two new datasets with coordinates from. The first dataset is the 'Parcs et jardins' database from Grand Paris Seine Ouest, that contains only publicly available parks and gardens in the Greater Paris region. Additionally, we use the 'Espaces verts' database from Mairie des Paris, containing a broader selection of green spaces for the inner city of Paris specifically. Both data sources are provided by the French government through www.data.gouv.fr.

Appendix II. Assumption testing

This appendix indicates whether our most complete model meets the assumptions of a linear regression. Furthermore, we describe the possible consequences for our results when the assumptions are not met. Table II.1 displays an overview of the technical notations and explanations of the five assumptions for linear regressions, as described by Brooks and Tsolacos (2010).

Table II.1 Assumptions linear regression (Brooks & Tsolacos, 2010)

Technical notation	Interpretation
1. $E(\varepsilon_t) = 0$	The errors have zero mean (linearity)
2. $V(\varepsilon_i) = \sigma^2 < \infty$	The variance of the errors is constant and finite over all values of x_i (homoscedasticity)
3. $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$	The errors are statistically independent of one another (no autocorrelation)
4. $\text{Cov}(\varepsilon_t, x_t) = 0$	There is no relationship between the error term and corresponding x variable
5. $\varepsilon_t \sim N(0, \sigma^2)$	ε_t is normally distributed

$E(\varepsilon_t) = 0$ – Linearity

The plot in figure II.1 indicates to what extent the residuals meet the predicted values. The average value of the errors should have a zero mean in order to meet this assumption. With this plot we find that our regression does not meet the assumption for linearity. However, Brooks and Tsolacos (2010) state that this assumption can never be violated when a constant is included, which is the case in the estimation.

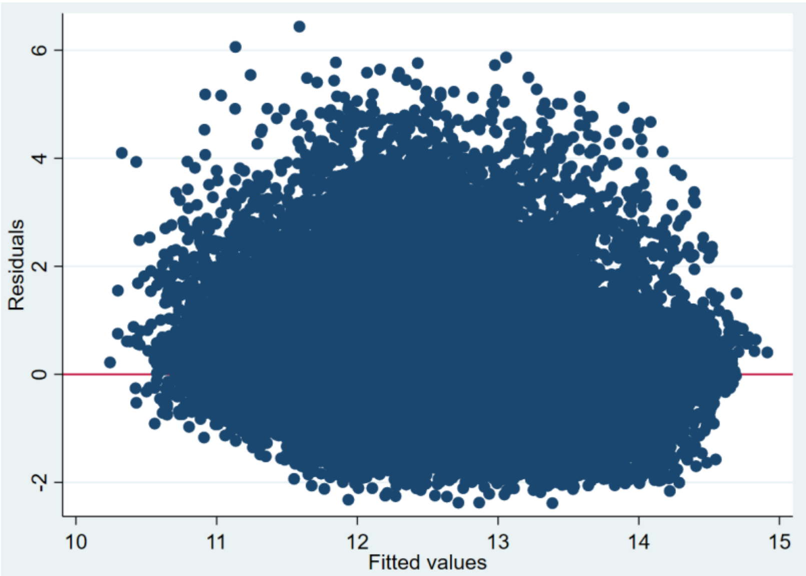

Figure II.1 Residuals plot OLS regression

$V(\varepsilon_i) = \sigma^2_{Y \cdot X}$ for $i = \infty$ – Homoscedasticity

The second assumption for the linear regression is homoscedasticity, where the variance of the errors should be constant. We assess the homoscedasticity with a Breusch-Pagan test, of which the results indicate that we should reject the null hypothesis. Therefore, the null hypothesis of constant variance for the error term is rejected. With heteroscedasticity, the coefficients are still unbiased. However, the standard errors cannot be relied on (Brooks and Tsolacos, 2010). In order to account for heteroscedasticity, we use robust standard errors in our regressions.

Cov(ε_i, ε_j) = 0 – no autocorrelation

Considering that data includes several years, autocorrelation is a potential problem for our regressions. In order to assess the presence of autocorrelation in our regression, we use both the Durbin-Watson test and the Breusch-Godfrey test. The Durbin-Watson statistic is approximately 1.95, which indicates that there is no autocorrelation. However, the estimations of the Breusch-Godfrey test indicate that we have to reject the null hypothesis of no serial autocorrelation. The consequences of autocorrelation are similar to the consequences of heteroscedasticity, which means that there is again an efficiency problem (Brooks and Tsolacos, 2010). However, we account for a potential autocorrelation problem with robust standard errors in our regressions.

Cov(ε_t, x_t) = 0 – no endogeneity

The presence of positive correlation between one or more explanatory variables and the error term will result in inconsistent coefficients. Therefore, in the case of endogenous variables, there is a problem with the interpretation of the coefficients (Brooks and Tsolacos, 2010). First of all, due to data limitations, we are not able to include all desired variables related to the quality of the house. Variables that reflect the quality of the house influence house prices, but can also be used by investors to assess investment opportunities. Accordingly, when these variables are both correlated with house prices and investment value, the coefficients of our investment variables are potentially biased. However, the investments relate to purchases in the close vicinity of the house, meaning that the properties bought by investors do not necessarily have similar housing characteristics.

Second, it is possible that house price levels in an area do affect investments as well. However, investors assess investment opportunities by considering the expected growth of house prices and not necessarily house price levels. Therefore, we do not expect endogeneity to be an issue, as we use house price levels as our dependent variable. This is also consistent with previous literature, where Lambie-Hanson et al. (2019) and Mills et al. (2019) use an instrumental variable approach to estimate house price growth, while Allen et al. (2017) does not consider an endogeneity problem with the estimation of house price levels. Nevertheless, it should be noted that house price growth rates are usually higher when house prices levels are lower, which implies that there is a relationship between house price levels and house price growth. However, due to our time-independent investment variables, we reveal correlations instead of causal effects with our model. Therefore, the influence of house price levels on investment values in the close vicinity is not considered a problem for our specific estimation anymore and we do not estimate a model with an instrumental variable approach.

$\boldsymbol{\varepsilon_t \approx N(0, \sigma^2)}$ – normality

Lastly, we test for normality, implying that the error terms are normally distributed or approaching a normal distribution. This assumption is required to make statements about the hypotheses.

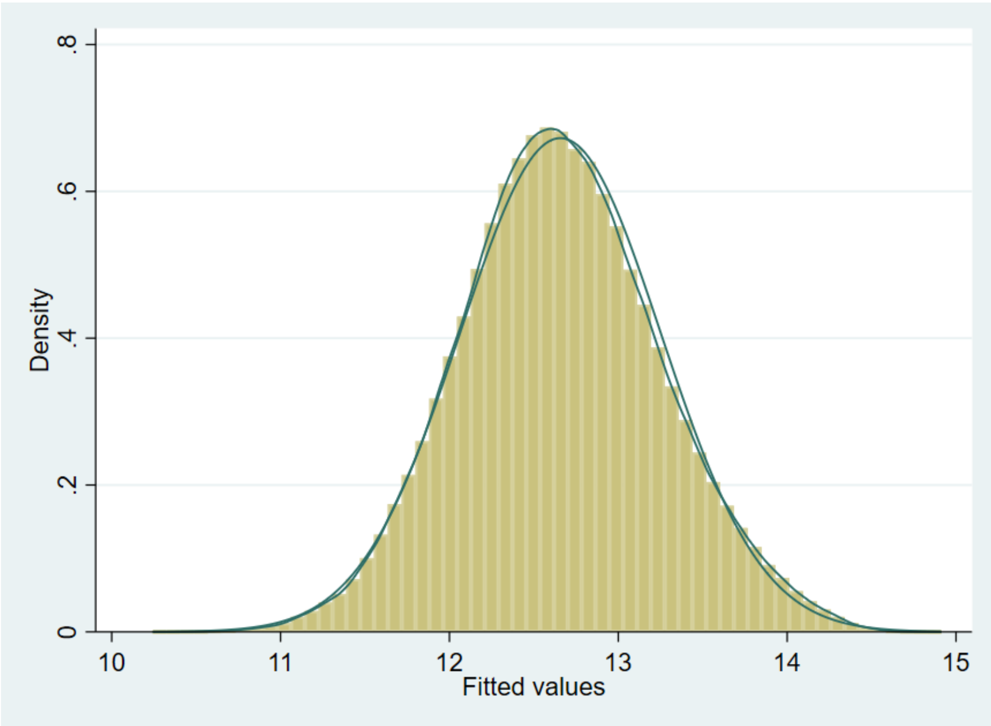

Figure II.2. Histogram residuals OLS regression

According to the histogram in figure II.2, the residuals seem normally distributed. However, the p-value of the Jarque-Bera normality test is lower than 0.005. Therefore, the null hypothesis of normal distribution of the error terms is rejected. Although normality is required to make inferences of the model, with a large dataset the consequences of not having a normal distribution are virtually inconsequential (Brooks & Tsolacos, 2010). Considering that we have a dataset with over 500,000 observations, we are still able to interpret the model.

Appendix III. Descriptive statistics heterogeneity groups

Apartments and single-family homes

Table III.1. Descriptive statistics single-family homes

Variable	Mean	Std. Dev.	Min	Max
transaction value	477214.25	826851.08	30000	1.140e+08
investment value radius 1500 m	21.108	50.792	0	649.3
investment value radius 2000 m	39.206	80.259	0	926.083
investment value radius 2500 m	63.459	116.275	0	1090.821
investment value radius 1500-2000 m	18.098	43.965	0	616.364
investment value radius 2000-2500 m	24.253	53.744	0	814.582
surface area of the house (m^2)	89.152	33.511	10	181
number rooms	4.071	1.23	1	7
single-family home (dummy, yes=1)	1	0	1	1
sale year	2017.077	1.937	2014	2020
distance closest public transport stop (km)	.422	.818	.001	8.045
distance closest park (km)	4.475	3.178	.003	14.819
Number of observations 72,290				

Table III.2 Descriptive statistics apartments

Variable	Mean	Std. Dev.	Min	Max
transaction value	431736.26	1241857.6	15000	1.655e+08
investment value radius 1500 m	124.506	134.22	0	728.818
investment value radius 2000 m	221.919	206.035	0	957.208
investment value radius 2500 m	343.914	294.299	0	1250.046
investment value radius 1500-2000 m	97.414	111.922	0	666.364
investment value radius 2000-2500 m	121.994	132.1	0	869.135
surface area of the house (m^2)	53.768	28.179	10	181
number rooms	2.521	1.143	1	7
single-family home (dummy, yes=1)	0	0	0	0
sale year	2017.056	1.945	2014	2020
distance closest public transport stop (km)	.139	.244	.001	7.534
distance closest park (km)	1.324	2.04	0	14.142
Number of observations 458,649				

Districts

Table III.3 Descriptive statistics district Paris

Variable	Mean	Std. Dev.	Min	Max
transaction value	570845.8	1597765.9	15000	1.506e+08
investment value radius 1500 m	217.189	132.562	0	728.818
investment value radius 2000 m	377.502	171.424	25.366	957.208
investment value radius 2500 m	574.944	212.467	81.792	1250.046
investment value radius 1500-2000 m	160.314	114.981	0	666.364
investment value radius 2000-2500 m	197.441	129.779	0	869.135
surface area of the house (m^2)	50.879	31.767	10	181
number rooms	2.352	1.173	1	7
single-family home (dummy, yes=1)	.004	.063	0	1
sale year	2017.025	1.962	2014	2020
distance closest public transport stop (km)	.1	.051	.002	.332
distance closest park (km)	.102	.068	.001	.53
Number of observations 207,445				

Table III.4 Descriptive statistics district Hauts-de-Seine

Variable	Mean	Std. Dev.	Min	Max
transaction value	436892.52	848480.47	18500	1.655e+08
investment value radius 1500 m	67.208	83.701	0	509.387
investment value radius 2000 m	130.594	145.952	0	673.26
investment value radius 2500 m	213.718	230.596	0	1131.282
investment value radius 1500-2000 m	63.386	86.165	0	610.109
investment value radius 2000-2500 m	83.124	109.782	0	828.417
surface area of the house (m^2)	63.109	31.203	10	181
number rooms	2.885	1.275	1	7
single-family home (dummy, yes=1)	.122	.327	0	1
sale year	2016.954	1.894	2014	2020
distance closest public transport stop (km)	.124	.084	.001	1.176
distance closest park (km)	1.664	1.557	0	6.906
Number of observations 127,406				

Table III.5 Descriptive statistics district Seine-Saint-Denis

Variable	Mean	Std. Dev.	Min	Max
transaction value	262536.91	836079.43	15500	1.140e+08
investment value radius 1500 m	30.803	65.353	0	485.72
investment value radius 2000 m	58.154	108.647	0	681.389
investment value radius 2500 m	98.488	159.969	0	738.147
investment value radius 1500-2000 m	27.351	63.485	0	500.405
investment value radius 2000-2500 m	40.334	81.76	0	520.999
surface area of the house (m^2)	62.554	28.684	10	181
number rooms	3.01	1.265	1	7
single-family home (dummy, yes=1)	.331	.47	0	1
sale year	2017.155	1.955	2014	2020
distance closest public transport stop (km)	.242	.291	.003	2.358
distance closest park (km)	4.369	3.155	.007	13.2
Number of observations 90,496				

Table III.6 Descriptive statistics district Val-de-Marne

Variable	Mean	Std. Dev.	Min	Max
transaction value	328366.23	782249.07	20000	1.327e+08
investment value radius 1500 m	21.074	40.473	0	261.014
investment value radius 2000 m	41.719	70.899	0	371.271
investment value radius 2500 m	65.462	100.316	0	480.216
investment value radius 1500-2000 m	20.645	48.835	0	356.259
investment value radius 2000-2500 m	23.743	50.813	0	371.902
surface area of the house (m^2)	64.868	30.025	10	181
number rooms	3.057	1.27	1	7
single-family home (dummy, yes=1)	.246	.431	0	1
sale year	2017.167	1.951	2014	2020
distance closest public transport stop (km)	.339	.796	.001	8.045
distance closest park (km)	2.863	2.633	.009	14.819
Number of observations 105,592				