DO YOUR NEIGHBOR'S LOOKS MATTER?

Aesthetic Appeal of the Urban Environment and House Prices: A Case Study of Lyon and Toulouse

JW van der Kam

12-02-2020

COLOFON

Title	**Do your neighbor's looks matter? Aesthetic appeal of the urban environment and house prices: a case study of Lyon and Toulouse**
Version	**Final**
Author	**JW van der Kam**
University	**University of Groningen**
Faculty	**Faculty of Spatial Sciences**
Student number	**S2914352**
E-mail	**jwvanderkam@gmail.com**
Supervisor	**dr. M.N. Daams**
Second assessor	**dr. M. van Duijn**

ABSTRACT

This study investigates the association between the aesthetic appeal of the urban environment and the prices of neighboring houses. Estimations are made by means of hedonic pricing models for two French cities: Lyon and Toulouse. A proxy for the aesthetic appeal of the urban environment is generated with the number of geo-located user-generated public images uploaded on online photo-sharing website Flickr. The findings show that the aesthetic appeal of the urban environment does have a positive association with nearby house prices. The results state that houses with several pictures within 50 meters are associated with at least 3.95 to 5.25 percent higher house prices, as opposed to houses with zero pictures within a 50-meter radius. The findings of this study can inform policy makers who make decisions that involve the regulation of aesthetics in the urban environment, maximizing well-being for urban citizens.

Key words: real estate, aesthetics, urban environment, hedonic price analysis, valuation

TABLE OF CONTENTS

1. INTRODUCTION

To protect the aesthetic appeal of the urban environment, most European countries, the United States, Australia, Canada and other countries have aesthetics committees in place to criticize building plans, of new constructions as well as renovations, in terms of spatial quality aspects (Federatie Ruimtelijke Kwaliteit, 2016). In this way, an urban environment that is aesthetically appealing for the majority of people is ensured. However, individuals or groups are sometimes not agreeing with the imposed rules. For example, a house owner in The Netherlands was obligated to repaint her bright green house in a more moderate color and went to court over it (Trouw, 2019). Furthermore, creative and innovative constructions that can be very beautiful do not fit within the given set of rules, which is limiting the overall aesthetic appeal. Discussions such as these have resulted in Dutch municipalities experimenting with removing aesthetics committees in their entirety or partially (Federatie Ruimtelijke Kwaliteit, 2016; Trouw, 2019; ZIN, 2019). Arguments used are that creative and outstanding buildings contribute to the city, that citizens will be able to take responsibility for their buildings and that the process of developing buildings will be shortened due to the reduction of procedures (Gemeente Gouda, 2004; Trouw, 2019). However, interests are different and some people just do not bother about the looks of what developers build (Trouw, 2011). Aesthetic committee free zones have resulted in many discussions, as people are free to do what they want to their properties in terms of appearance (Archined, 2004; Trouw, 2019; Trouw, 2011). As a result, neighboring property owners are afraid that their properties will decrease in value due to untasteful designs in the near vicinity (Trouw, 2011).

Some countries, such as France and Belgium, require architects to be involved in the designs of most new constructions and renovations (Droit-Finances, 2019; Federale Overheidsdienst Economie, 2019a). The title *"architect"* is protected and can only be used after the completion of a specific educational degree plus several years of relevant work experience (Federale Overheidsdienst Economie, 2019b; JUSTEL, 2019). By requiring an architect, these countries try to protect the aesthetic appeal of the urban environment. However, Belgian houses are often belittled as they can be very distasteful, to the point of entire blogs and books that are dedicated to it (Caudenys, 2015; Bouw Wereld, 2011).

Location has a large influence on buyers' willingness to pay for residential properties. People want to live in places that they appreciate for varied reasons. Next to the specific country, state, city or neighborhood of preference in which the dwelling is located, factors such as air quality, public green space and waterfronts all have their proven influence. When zooming in a bit more, the available amenities of the urban environment, such as the proximity of a supermarket, shopping center, gym or church, are all factors that buyers of property value and take into account in their willingness to pay. Another plausible factor could be the aesthetic appeal of the direct urban environment.

This research focuses on the association between the aesthetic appeal of the direct urban environment and nearby house prices in two French cities. An economic problem arises when aesthetics enter housing prices only for occupiers, while they fail to account for externalities imposed on their neighbors. When there is an influence of aesthetics beyond one's own home outside of the price system, it is likely that policy should account for this. The question is how these implicit prices of aesthetic design can be estimated in a correct way, in order to not underprovide or overprovide for the effects. The results of this study can inform urban planners, policy makers and scientists looking to explain the social and economic outcomes of urban development.

There have been several studies that estimated, for example, the effect of historic, monumental and iconic buildings on nearby house prices. Most of these studies are specifically focused on one type of "special" building, and not so much the broader category of beautiful buildings in the vicinity. What all studies, however, do have in common is that positive effects are found. More information on these studies can be found in the literature review and theory section.

A recent study of Saiz, Salazar and Bernard (2018) found that buildings with higher beauty ratings are more likely to be geotagged with user-uploaded photos on Google Maps and Flickr. This finding validates the use of localized user-generated image uploads on photo-sharing websites to measure the aesthetic appeal of the urban environment (Saiz, Salazar and Bernard, 2018). The number of photos will serve as a proxy variable. A proxy variable is defined as a variable that is not in itself directly relevant, but serves in place of an unobservable or immeasurable variable (Upton and Cook, 2008). Seresinhe, Moat and Preis (2018) found that models with data from Flickr can generate more accurate estimates on how scenic people consider an area to be than models that consider only basic census measurements. That Flickr can serve as a reliable, universal source of spatial content is showed by Antoniou *et al.* (2010). There is a growing number of studies that have used content from photo-sharing websites. Some examples are Ahlfeldt (2013), who used photo upload-frequencies to identify high-amenity areas in Berlin and London and Paldino *et al.* (2015), who were able to show that tourists and native city-dwellers display similar photo uploading patterns.

In summary, a number of studies estimate the effects of for instance historic buildings, monumental buildings, iconic buildings or buildings of a specific building style. However, most of these studies are specifically focused on one type of particular building and not so much the extensive category of beautiful buildings in the surrounding area. What is missing is a broader study that estimates the association between beautiful design and nearby houses. Following Saiz, Salazar and Bernard (2018), who validated the use of localized user-generated image uploads on photo-sharing websites as a proxy for the aesthetic appeal of the urban environment, it becomes a possibility to estimate the association between house prices and nearby beautiful buildings of that broader category. Daams, Sijtsma and van der Vlist (2016) stress the importance of informing hedonic price models about perceptions of location quality. The validation of the proxy for aesthetics in the urban environment, by Saiz, Salazar and Bernard (2018) included respondents ranking the buildings in the study according to their own perceptions. By using user-generated image uploads as a proxy, this study accommodates Palmquist (2005), who states that measures used in hedonic price analysis should capture as precisely as possible the way potential property buyers perceive an observed environmental good. In specific, the research aim of this study is to estimate the association between the aesthetic appeal of the urban environment and nearby house prices, using geotagged user-generated image uploads of a photo-sharing website. The central research question for this study is:

Is there an association between the aesthetic appeal of the adjacent urban environment and house prices?

The association between the aesthetic appeal of the urban environment and nearby house prices is analyzed with property transaction data and photo-frequency data of two French cities, Lyon and Toulouse, with a hedonic pricing model. Usable data for 56 percent of property transactions in these two cities is obtained. Lyon and Toulouse are chosen for their characteristics in terms of population and location. More information on this choice is given in the data and methodology section of this thesis.

This thesis has several sections as follows: the next section, the literature review and theory section, will summarize relevant literature and describe the most important concepts that are relevant for this study. Section 3, the data and methodology section, describes the data sources, selection criteria, descriptive statistics, the research method and the way that the analysis has been done. Section 4, the results section, which presents the results of the empirical analysis. Section 5, the discussion section, in which the main findings will be discussed. The final chapter will give an overall conclusion.

2. LITERATURE REVIEW & THEORY

In trading theory, it is often assumed that products are heterogenous and are being traded on transparent markets with information symmetry. According to Tiwari and White (2010), these theories do not apply to real estate as real estate is very heterogenous, which makes different real estate objects difficult to compare. There are numerous differences in real estate objects, such as the building materials, the number of rooms and the size of the garden. Next to these physical characteristics, the price of real estate is very dependent on the location, as it cannot be moved. Montero, Fernández-Avilés and Mínguez (2018) concluded that the environment has a significant effect on house prices. Certain locations are more wanted than others, and this is reflected in the prices of houses (DiPasquale and Wheaton, 1996; Alonso, 1960). Because of all these differing characteristics, real estate markets are not transparent and buyers do not enjoy full information on the market (Tiwari and White, 2010). Due to the inefficiency of the market, it is not always clear which factors contribute with what magnitude to real estate prices. As a means to investigate these contributions, Rosen (1974) introduced a hedonic price model that explains differences in property value to characteristics featured in the properties using regression analysis. Hedonic price modeling is often used to estimate property demand and property values with respect to certain amenities, either in or around the property (Celik and Yankaya, 2006).

The importance of location in relation to residential property values is a well-researched topic in property valuation literature. The literature tells that amenities create external effects that are reflected into property values (Cheshire and Sheppard, 1995; Wilkinson, 1973). There are many examples to be found for studies that examined the effects of external amenities on house prices, nearly all relying on a hedonic price analysis. For instance, Daams, Sijtsma and van der Vlist (2016) estimated the effect of attractive natural space on Dutch residential property prices, Han (2014) studied the impact of abandoned properties on nearby property values in Baltimore (United States) and Bae, Jun and Park (2003) investigated the impact of a new subway line on residential property values in Seoul (South Korea).

As mentioned in the introduction, there have been several studies that estimated the effect of historic, monumental and iconic buildings on nearby house prices. Most of these studies are specifically focused on one type of "special" building, and not so much the broader category of beautiful buildings in the vicinity. Ruijgrok (2006) was able to show that historical characteristics of buildings and their surroundings account for almost 15 percent of property values by studying buildings in Tiel, the Netherlands. She stated that this is mainly due to the authenticity and appearance of the buildings (Ruijgrok, 2006), therefore it may apply to all authentic and beautiful architecture and their surrounding environment. Lazrak, Nijkamp, Rietveld and Rauwendal (2014) found that buyers in the Zaanstad urban area are willing to pay an additional 0.28 percent for each additional listed heritage building that lies within a 50-meter radius, proving that cultural heritage is reflected in house prices as well. According to a study of Levi (2016) both real and fake historic architecture is perceived as attractive by people, even though they are capable of discriminating between them. In terms of iconic buildings, Ahlfeldt and Mastro (2012) conducted a case study on the iconic design of 24 residential structures by world-famous architect Frank Lloyd Wright. A premium of 5 percent was found within a 50-meter distance, indicating that an external premium to iconic architecture does exist (Ahlfeldt and Mastro, 2012). Historic, monumental and iconic buildings are in general assumed to be aesthetically appealing and therefore contributing to the aesthetic appeal of the urban environment.

A qualitative study executed by Millhouse (2005) assessed the effect of architectural design on real estate values through interviews with industry leaders and policy makers, and a literature study. He concludes that good design does not only have an effect on the property in question, but also indirectly benefits the surrounding property values because good design can be used as a marketing device, increasing absorption and decreasing vacancy nearby (Millhouse, 2005). Song and Knaap (2003) studied the virtues of new urbanism, a very significant movement in urban planning and architecture. They concluded that there is a measurable value difference between traditional and new urbanist neighborhoods that can be capitalized into residential property values (Song and Knaap, 2003). On the basis of their measures of design character of urban neighborhoods, their results show that the price premiums, or discounts, of any particular neighborhood in Washington County (Oregon, United States) depends on the particular design characteristics it has to offer (Song and Knaap, 2003).

The aesthetic appeal of the urban environment can be described as a view towards buildings in the direct environment (Bourassa, Hoesli and Sun, 2004). Therefore, the aesthetic appeal of the urban environment has the same underlying concept as the effect of a view and the value that people award to a view. The views that can be seen from a residential unit are generally regarded to have a positive impact on the value of the property (Ming and Hian, 2005; Darling, 1973; Bond, Seiler and Seiler, 2002; Plattner and Campbell, 1978; Rodriguez and Sirmans, 1994; Benson et al., 1998; Bourassa, Hoesli and Sun, 2005; Nicholls and Crompton, 2018). Ming and Hian (2005) were even able to show that an obstruction of views depresses the prices of the obstructed development by 8 percent in the long run. Bourassa, Hoesli and Sun (2004) concluded by a study in Auckland, New Zealand, that a view on attractive buildings in the neighborhood of a property on average adds 37 percent to value relative to properties in neighborhoods with only average-quality structures.

Brueckner, Zenou and Thisse (1999) presented an amenity-based theory of location by income. Their theory shows that the relative location of different income groups depend on the spatial pattern of amenities in a city (Brueckner *et al*., 1999). The marginal valuation of amenities rises sharply with income, therefore the higher income groups go live where the amenities are (Brueckner *et al.*, 1999). Brueckner *et al*. (1999) classified the amenities into three categories: natural amenities (e.g. rivers and forests), historical amenities (e.g. monuments and buildings) and modern amenities (e.g. restaurants and theaters). The historical amenities can be assumed to highly correlate with the aesthetic appeal of the urban environment. Especially the preferences for exogenous amenities (natural and historical amenities) determine the location of the higher income groups (Brueckner *et al*., 1999). Modern amenities, such as restaurants and theaters, tend to follow the higher income groups and thus come up where they choose to live (Brueckner *et al*., 1999). European cities, including Lyon and Toulouse, have many historical amenities, resulting in the higher income groups wanting to live near those amenities (Brueckner *et al*., 1999). Adding to this, Glaeser, Kolko and Saiz (2001) stated that higher educated people (with usually a higher income) often live in areas with high amenities. As a result, prices of houses near amenities (such as the aesthetic appeal of the environment) should be higher as opposed to prices of houses far away from amenities, as the higher income groups have a higher willingness to pay due to their higher income. Demand and supply mechanisms will assure that there is a certain value attached to certain amenities (Besanko *et al*., 2013). It should be noted that not only the higher income groups will seek to live near amenities. Any individual of any income group values different amenities differently. Each individual will seek to maximize their utility at a given income (Rosen, 1974).

Many characteristics of houses and their environment are reflected in the price buyers are willing to pay. The aesthetic appeal of the urban environment can be seen as an amenity. Based on the amenity-based theory of Brueckner *et al*. (1999), higher income groups are willing to pay to live near amenities, driving up real estate prices. Therefore, it is expected that the aesthetic appeal of the urban environment has a positive association with nearby house prices. Historic, monumental and iconic buildings are in general assumed to be aesthetically appealing and therefore contributing to the aesthetic appeal of the urban environment. Consequently, examples of earlier studies in the literature review on historic, monumental and iconic buildings and their positive effects on house prices also feed the expectation that the aesthetic appeal of the urban environment is associated with higher real estate values. The following hypothesis is tested in an empirical way:

The aesthetic appeal of the urban environment is associated with nearby house prices

3. DATA & METHODOLOGY

3.1. User-generated photo data

For this study, different types of data are required. As mentioned in the introduction, Saiz, Salazar and Bernard (2018) validated the use of localized user-generated image uploads on photo-sharing websites to measure the aesthetic appeal of the urban environment. For background information on the validation of the proxy, see the paper of Saiz, Salazar and Bernard (2018). Photo data are obtained from Flickr by the use of Flickr API. The experimental plugin called *Flickr Metadata Downloader* enables the extraction of geotagged public Flickr photos within a given geographic quadrangle through QGIS (Quantum GIS), an open source Geographic Information System. The extracted metadata contains the geographic position, photo ID, date, time, accuracy level, photo title, tags and the URL of a small thumbnail image. Downloading the Flickr metadata has proven to be a timely process, forcing the user of the QGIS plugin to reevaluate the size of the study area. For this study, photo metadata of Lyon and Toulouse is obtained. These are the only two French landlocked cities, except from the capital Paris, that are in the top five of largest French cities in terms of population. Land locked cities were chosen to avoid the enormous number of geotagged pictures along the coastline in seaside cities, such as Marseille and Nice, that could interfere with the results. For Lyon, metadata for a number of 148,442 photos were obtained. For Toulouse, metadata for a number of 96,472 photos were obtained. The total extracted number of photo information observations is 244,914. These photos were all taken and uploaded before the 16th of October 2019, which was the moment of data extraction.

3.2. Property transaction data

Property transaction data of France is used for the empirical analysis. These data are publicly available through Etalab on the www.data.gouv.fr data platform. Etalab is a French government agency that engages itself with the coordination of the policy of openness by means of sharing open data (Etalab, 2019). Specifically, the *demandes de valeurs foncieres géolocalisées* datasets from 2014 till 2018 are used. These datasets contain geolocated transaction values. The combined datasets for the whole country of France for these five years include data on 13,903,117 real estate transactions. As described in paragraph 3.1, Lyon and Toulouse were selected to be examined for this study by reason of time limitations and their features in terms of location and population size. The property transaction data for Lyon and Toulouse contain respectively 87,876 and 121,630 observations of object transactions from 2014 till 2018. This totals up to 209,506 observations of real estate transactions. Transaction value will be the dependent variable in the analysis. Further, the datasets contain more independent variables that give information on transaction characteristics, location characteristics and building characteristics. The following variables are found to be useful for the analysis: mutation date, transaction value, postal code, commune code, type of building, surface area and number of main rooms. With the use of the mutation date variable, a transaction year variable could be created. The postal code variable is used to check for location fixed effects, which controls for omitted variable bias. In total, 15 postal codes are present in Lyon and Toulouse. A limitation of the dataset is that it does only contain a few building characteristics. More characteristics, such as information on maintenance, central heating, outside space, building period and parking possibilities would have been useful to increase the predicting power of the empirical model.

By combining the property transaction data with the photo metadata, specifically the geographic position, a variable with the number of photos in a radius of 50 meters around each transacted property was created with the use of Geographic Information Systems (GIS). A radius of 50 meters was chosen because Saiz, Salazar and Bernard (2018) state that the correlation between image uploads and building beauty vanishes for photos taken more than 50 meters from the building of interest. Therefore, in general, photos that are further than 50 meters away are not contributing to the aesthetic perception and will not be used to create scalable quantitative measures for aesthetic perception. Saiz, Salazar and Bernard (2018) were aware of the many factors affecting the propensity of internet users to upload photos, such as the iconic status of buildings or their location in touristic areas. Random noise in the data, amounted by for example the aforementioned reasons, was proven to not invalidate that the photo-uploading frequency predicts average subjective beauty ratings (Saiz, Salazar and Bernard, 2018).

There is a possibility that there are variables that are not included in the dataset, and therefore not included in the model, that are associated with house prices. This would result in omitted variable bias. The consequence would be that the estimated coefficients on all the other variables will be biased and inconsistent, standard errors would be biased, the estimate of the coefficient on the constant term would be biased and hypothesis tests could bring forth inappropriate inferences (Brooks and Tsolacos, 2010). By means of spatial controls and controls for time fixed effects, this will be controlled for as much as possible. However, due to data and time limitations this study will be conducted with the variables that are available in the datasets mentioned above.

3.3. Data selection

Unfortunately, not all previously described data is seen fit for the analysis. The metadata of the Flickr photos contain the accuracy level variable. Before the combining of the two datasets the following selection took place. Only data points with the maximum accuracy level (16), which corresponds to photos which already contain GPS coordinates from the camera or those where the user has zoomed into the relevant street in order to pin down the photo to its location, are maintained. Photos with a lower accuracy level are not usable as it cannot be surely said that they lay within a 50-meter radius of a certain building. This measure resulted in 58,002 dropped photo locations from Lyon and 48,434 dropped photo locations from Toulouse. Next to this, a total number of 211 photo locations were deleted due to them being from before the year 2000. What is left is a total number of 138,267 usable photo locations.

In the combined dataset, there is a number of 209,506 real estate transactions which have a number of photos assigned to each of them. However, the dataset is not perfect, resulting in the fact that a full sample study is not possible. Mutations that went through in one go, all separately display the transaction value of the whole transaction. This might be a house with an outbuilding (shed or barn that belongs to a main building), in which the outbuilding is separately listed for the price of the house as well. For this reason, all observations of outbuildings are dropped (65,863 observations). Furthermore, large transaction deals state the transaction price for the whole deal (e.g. 80 apartments that are sold by one seller to one buyer). This leads to separate houses being listed for millions of euros in the dataset. Therefore, all duplicates in terms of mutation ID are dropped (including the first duplicate) (62,598 observations) to ensure the validity of the dataset. The reason that the outbuildings are removed before the duplicates are dropped, is that all observations of individual houses with an outbuilding would get lost otherwise. What remains is 56 percent of all property transactions in Lyon and Toulouse. Next, all observations of industrial and commercial buildings are removed (3,314 observations). All observations without any building type information are removed as well (3,297 observations).

In terms of transaction price, all observations below the 1^{st} percentile and above the 99^{th} percentile are removed (1500 observations). Incomplete observations in terms of coordinates (505 observations), postal code (53 observations) or number of rooms (328 observations) were removed as well. Furthermore, outliers in terms of surface (below 17 square meters, below 1^{st} percentile) (661 observations)and main rooms (above 6 rooms, above 99^{th} percentile) (336 observations) are deleted. With all above selection criteria applied, a total number of 71,051 observations is left.

3.4. Descriptive statistics

In table 1 the descriptive statistics on the remaining 71,051 observations are given. They are divided into transaction variables, building characteristic variables and photo variables. The transaction prices range from 35,000 to 720,100 euros, with a mean price of almost 195,000 euros and a standard deviation of approximately 113,000 euros. As described before, data from 2014 till 2018 is used. The mean indicates that there were more transactions in the earlier years. Further investigating the transaction year variable shows that there are significantly less transactions from the year 2018 in the dataset. Even though transactions from December 2018 are included in the dataset, it is plausible that not all transactions from the year 2018 were incorporated in the Etalab dataset for 2018 at the time of extraction (September 2019).

Regarding the surface area, the smallest house is 17 square meters while the largest house is 300 square meters. The mean surface area is approximately 62 square meters (standard deviation = 28). In terms of rooms, the houses in the dataset consist of 1 to 6 rooms, with a mean of 2.7 rooms (standard deviation = 1.18). The building type variable is categorical, in which the number 1 indicates a *maison* (includes detached houses, semi-detached houses and terraced houses) and the number 2 indicates an apartment. The mean of 1.928 strongly indicates that the majority of houses in Lyon and Toulouse are apartments, while a small portion of the houses is not an apartment.

The number of photos within 50 meters of an individual house ranges from 0 to 1,016, while the mean number of photos is 8.883. The standard deviation is relatively large at 37.072 photos. A number of 42,844 houses in the dataset does not have any photos in a 50-meter radius. This means that 28,206 houses do have at least one photo assigned to them, which is a percentage of 39,7 percent. The majority of those, 20,221 houses, have 1 to 10 photos within 50m. Lastly, the dummies for the photo classes are included in the descriptive statistics to emphasize that these will be used in the empirical analysis. The first five photo classes variables will not be included in the model and are solely to give insight in the distribution of photos among the houses. For the analysis, the transaction price and surface area variables were transformed into natural logarithms to ensure they are closer to a normal distribution. This can be seen in Appendix A.

Table 1: Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Transaction price	71051	193932.2	112775.9	35000	720100
Transaction year	71051	2015.841	1.295	2014	2018
Surface area	71051	61.795	28.171	17	300
Rooms	71051	2.740	1.185	1	6
Building type	71051	1.928	0.258	1	2
Number of photos <50m	71051	8.883	37.072	0	1016
Photos 0	42844	0	0	0	0
Photos 1-10	20221	3.076	2.466	1	10
Photos 11-20	2191	14.936	2.932	11	20
Photos 21-30	1249	25.400	2.934	21	30
Photos 31+	4546	110.985	100.184	31	1016
D Photos 0	71051	.603	.489	0	1
D Photos 1-10	71051	.285	.451	0	1
D Photos 11-20	71051	.031	.173	0	1
D Photos 21-30	71051	.018	.131	0	1
D Photos 31+	71051	.064	.245	0	1

3.5. Empirical models

This study will use hedonic models to estimate the association between the aesthetic appeal of the urban environment and nearby house prices. The hedonic models will be estimated using ordinary least squares (OLS). OLS makes it possible to estimate the relationship between the dependent and the independent variables. For a detailed discussion on the testing of the OLS assumptions and the assessment of multicollinearity, see appendices B and C. The dependent variable in this study is the natural logarithm of the transaction price. As independent variables, several building, location and transaction characteristics are included. The hedonic framework will be as follows:

$$P = f\ (B, L, T) \tag{0}$$

In this formula, P represents the house price, that is determined by different factors (*f*). The factors are building characteristics (*B*), location characteristics (*L*) and transaction characteristics (*T*). Building characteristics available are surface area, number of rooms and type of house. Location characteristics available are the postal code and the number of photos within 50 meters. Transaction year is a transaction characteristic. Several models are specified. First of all, the baseline model in which all available building, location and transaction variables are included except for the photo-variables:

$$\log(Price) = \beta_0 + \beta_1 \log(Surface_i) + \beta_2 BuildingType_i + \beta_3 \sum_{j=2}^{6} Rooms_{ij} + \beta_4 \sum_{k=2}^{15} PostalCode_{ik} + \beta_5 \sum_{m=2}^{5} Year_{im} + \varepsilon_i \tag{1}$$

In this formula, and the following formulas, *i* indicates the specific property. *BuildingType* is a dummy variable indicating the building type (out of two possible types). *Rooms* is a vector for dummies for each number of rooms (*j*). *PostalCode* is a vector for dummies for each postal code (*k*). The postal code variable is used for location fixed effects, which controls for omitted variable bias. In total, 15 postal codes are present in Lyon and Toulouse. *Year* is a vector of dummies for each year (*m*). The year variable is used for time-fixed effects. Secondly, the number of photos within 50 meters variable (*Photos*) is added to the model. This gives the following specification:

$$\log(Price) = \beta_0 + \beta_1 Photos_i + \beta_2 \log(Surface_i) + \beta_3 BuildingType_i + \beta_4 \sum_{j=2}^{6} Rooms_{ij} + \beta_5 \sum_{k=2}^{15} PostalCode_{ik} + \beta_6 \sum_{m=2}^{5} Year_{im} + \varepsilon_i \tag{2}$$

Third, the specification which includes four classes of photos in order to observe possible differences among the associations for different numbers of images and therefore aesthetics of the environment:

$$\log(Price) = \beta_0 + \beta_1 Photos_{1-10,i} + \beta_2 Photos_{11-20,i} + \beta_3 Photos_{21-30,i} + \beta_4 Photos_{31+,i} + \beta_5 \log(Surface_i) + \beta_6 BuildingType_i + \beta_7 \sum_{j=2}^{6} Rooms_{ij} + \beta_8 \sum_{k=2}^{15} PostalCode_{ik} + \beta_9 \sum_{m=2}^{5} Year_{im} + \varepsilon_i \tag{3}$$

There is a possibility that the postal code variable partially absorbs the magnitude of the association of urban aesthetics and nearby house prices, as houses in the same postal code area may have similar aesthetic characteristics. For this reason, the previous two models are also examined without the postal code dummy-variables. This leads to the following two specifications:

$$\log(Price) = \beta_0 + \beta_1 Photos_i + \beta_2 \log(Surface_i) + \beta_3 BuildingType_i + \beta_4 \sum_{j=2}^{6} Rooms_{ij} + \beta_5 \sum_{m=2}^{5} Year_{im} + \varepsilon_i \tag{4}$$

$$\log(Price) = \beta_0 + \beta_1 Photos_{1-10,i} + \beta_2 Photos_{11-20,i} + \beta_3 Photos_{21-30,i} + \beta_4 Photos_{31+,i} + \beta_5 \log(Surface_i) + \beta_6 BuildingType_i + \beta_7 \sum_{j=2}^{6} Rooms_{ij} + \beta_8 \sum_{m=2}^{5} Year_{im} + \varepsilon_i \tag{5}$$

The variables of interest for this study are the photo variables that capture the aesthetic appeal of the urban environment. Further, in all five models, ε is the error term of the model. The betas (β) are the parameters that are to be estimated.

It is expected that the models will result in positive coefficients between the number of photos and nearby house prices. However, the magnitude and significance of these coefficients will differ between models. The last two models are expected to show positive associations with a higher magnitude, as potential absorption of the magnitude of the association by the postal code dummies is removed.

4. RESULTS

4.1. Results for models with spatial controls

The results for the different models are presented in table 2. Model (1) is the baseline specification. The adjusted R-squared value of the model is 0.650, which means that the model without any photo-variables explains 65 percent of the variance in house prices. Included variables in this model are building characteristics, location characteristics (spatial fixed effects) and transaction characteristics (year fixed effects in the case of this study). Significant coefficients for the natural logarithm of surface area and building type were expected, as a larger house in terms of surface area is found to sell for a higher price and apartments are generally shown to sell at lower prices than other housing types.

Models (2) and (3) have adjusted R-squared values of respectively 0.650 and 0.651, which means that the photo-variables do add very little explanatory power to the models and they still predict approximately 65 percent of the variance in house prices. The coefficients of the control variables remain the same as in model (1). In model (2), the number of photos within a 50-meter radius is the variable of interest. The coefficient for this variable is slightly positive but not significantly different from zero. In model (3), the number of photos within a 50-meter radius variable is replaced by several classes of photos within a 50-meter radius. Although all four classes have positive coefficients, only the coefficients of the first three classes are significant. The results show that, as opposed to having 0 photos in a 50-meter radius, having 1 to 10 photos in a 50-meter radius is associated with 3.95 percent[1] higher house prices, having 11 to 20 photos in a 50-meter radius is associated with 4.67 percent higher house prices and having 21 to 30 photos in a 50-meter radius is associated with 5.25 percent higher house prices. The remaining class of 31+ photos within a 50-meter radius has no significant coefficient and therefore, on the basis of this model, it can be concluded that there is no association for buildings with more than 30 images within a 50-meter radius. A Chow test confirmed that a pooled model with Lyon and Toulouse is valid.[2]

4.2. Results for models without spatial controls

In models (4) and (5), the postal code dummies are omitted because there is a possibility that they partially absorb the magnitude of associations between aesthetics in the environment and nearby house prices. Omitting the postal code dummies resulted in lower R-squared values (57 percent and 59

---

[1] Calculated as (exp(0.0387)-1)*100 following Halvordsen and Palmquist (1980)

[2] As a robustness analysis, a Chow test is performed to evaluate whether the parameters for the two different cities, Lyon and Toulouse, have equal coefficients. The null hypothesis that the true intercepts and slopes are identical between predefined subsamples cannot be rejected on the basis of this test (F = 0.66, p = 0.4305). Therefore, it can be concluded that there is no break in the dataset and a pooled model with the two cities is valid.

percent respectively), meaning the models have less explanatory power. The coefficients and significance levels of the photo variables change drastically in comparison to the earlier models, while those of the remaining control variables remain relatively similar. Model (4) shows that the coefficient of the number of photos within 50 meters variable is 0.00097, which is significant at the 1 percent level. This would mean that an increase of 1 photo within a 50-meter radius is associated with a 0.097 percent higher house price. Model (5) shows positive coefficients that are significant at the 1 percent level for all photo class variables. It shows that having 1 to 10 photos in a 50-meter radius is associated with a 16.07 percent higher house price, having 11 to 20 photos within 50 meters is associated with a 23.37 percent higher house price and having 21 to 30 photos within 50 meters is associated with a 26.11 percent higher house price. On top of that, the results on the previously insignificant class shows that having more than 30 photos within a 50-meter radius is associated with a 25.36 percent higher house price.

Table 2: OLS estimates of house price models

	(1) Log Transaction price	(2) Log Transaction price	(3) Log Transaction price	(4) Log Transaction price	(5) Log Transaction price
Number of photos <50m		0.000293		0.00097***	
		(0.000236)		(0.0000398)	
D Photos 1-10			0.0387**		0.1494***
			(0.0167)		(0.0030)
D Photos 11-20			0.0456**		0.210***
			(0.0192)		(0.0074)
D Photos 21-30			0.0512*		0.232***
			(0.0274)		(0.0098)
D Photos 31+			0.0468		0.226***
			(0.0459)		(0.0057)
Log Surface area	0.835***	0.834***	0.834***	0.990***	0.957***
	(0.0237)	(0.0236)	(0.0232)	(0.0067)	(0.0065)
Building type = 2	-0.333***	-0.334***	-0.337***	-0.171***	-0.214***
	(0.0406)	(0.0406)	(0.0403)	(0.0064)	(0.0064)
Constant	8.981***	8.983***	8.969***	8.352***	8.422***
	(0.0803)	(0.0799)	(0.0750)	(0.0237)	(0.0230)
Room dummies	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Spatial controls	Postal code	Postal code	Postal code	None	None
Observations	71,051	71,051	71,051	71,051	71,051
Adj. R-squared	0.650	0.650	0.651	0.570	0.590

Notes: The reference categories include 0 photos within 50m and houses (as opposed to apartments). D stands for Dummy. In parentheses are the standard errors, that are robust to heteroscedasticity for specifications (1) to (5) and clustered at the postal code level for specifications (1), (2) and (3). *, **, *** Significance at 10%, 5% and 1% respectively.

5. DISCUSSION

The results of this study show that the models derived significant, positive results on the association of the aesthetic appeal of the urban environment and nearby house prices. Recall for this that the number of geotagged user-generated image uploads within a 50-meter radius is a proxy for the aesthetic appeal of the urban environment (Saiz, Salazar and Bernard, 2018). In model (3) the number of photos variable was divided into multiple classes. From the results, it can be seen that only the highest class (31 and more photos) delivers no significant coefficients. The other three classes (1 to 10, 11 to 20 and 21 to 30 photos) do deliver significant coefficients. These lower three classes together, make up 81.9 percent of the observations (not including the reference category of 0 photos). Based on the average house price of 193,932 euros, the association translates to a value of 7,660 euros (3.95 percent) to 10,181 euros (5.25 percent), depending on the amount of geotagged user-generated image uploads in the near vicinity. In the urban environment, where houses are often built closely together, it may sum up to a very substantial value. It can be observed that the magnitude of the association strengthens when the number of photos increases. However, might the number of image uploads rise above 30, this model predicts that there is no association, as it does not provide significant coefficients on this class. It might be a possibility that the aesthetic appeal, expressed in 31 or more photos, does not have a positive association as other (negative) factors come into play, such as decreased privacy or lack of peace and quietness on the street. Another explanation would be that the aesthetic appeal of these locations goes beyond the 50-meter radius and is fully picked up by the postal code fixed effects. This would require further research. It should be noted that this class of 31 and more photos consists of only 6.4 percent of the total observations. When looking at the spatial distribution of the different photo classes, it can be seen that the 31 plus photo class displays a clustered pattern (see appendix D, red markers). The clusters of the class happen to be in the high-demand, high-amenity areas of Lyon and Toulouse. It might be a possibility that the unobserved quality of property comes into play in these locations.

To examine this further, in models (4) and (5), the postal code dummies are omitted from the models because they might partially absorb the magnitude of the association of aesthetics in the urban environment and nearby house prices. Model (4) shows that an increase of every 10 photos within a 50-meter radius is associated with a 0.97 percent higher house price. This translates to 1745 euros per 10 additional photos for the average house in the sample. Model (5) results in relatively large associations for all photo classes, ranging from 16.07 to 26.11 percent, equaling 31,164 to 50,636 euros of value for the average house. What can be observed is that the association increases in magnitude till a number of 30 pictures within 50 meters, and then declines a small bit (0.75 percent) for the rest of the observations. With the omission of the postal code dummies, however, it is very likely the case that the number of photos is absorbing more associations and effects than only the aesthetic appeal of the urban environment, as there are no longer spatial control measures present in the models. For this reason, the association of the aesthetic appeal of the urban environment and nearby house prices will most likely not be as substantial as outlined in model (5), though it is very plausible that the associations are bigger than outlined in model (3) and present for all photo classes. With these results, the hypothesis that the aesthetic appeal of the urban environment has a positive association with nearby house prices cannot be rejected. There is a possibility that the association is reflecting a causal effect, meaning the aesthetic appeal of the urban environment has a direct influence on nearby house prices. This can, however, not be certainly said on the basis of this study. For this reason, an association is claimed instead of a causal effect. Knowledge of an association is the first step to proving a causal relationship between the two concepts.

The positive results are in line with previous studies on historic (Ruijgrok, 2006), monumental (Lazrak *et al*., 2014) and iconic buildings (Ahlfeldt and Mastro, 2012), which are generally assumed to be aesthetically appealing and therefore contributing to the aesthetic appeal of the urban environment. This study quantitatively confirmed the positive association that the broader class of beautiful buildings has with nearby house prices, confirming the hypothesis of Millhouse (2005) in his qualitative study. The outcome of Bourassa, Hoesli and Sun (2004), that a view on attractive buildings in the neighborhood of a property adds on average 37 percent to value relative to properties in neighborhoods with only average-quality structures, is not nearly met. This is presumably as Bourassa, Hoesli and Sun (2004) estimated the effect of superior views on attractive buildings, where this study filtered out outliers. This study adds to the growing number of studies using public user-generated image data and can serve as an inspiration for future research.

The results of this study can be useful for urban planners and scientists looking to explain the social outcomes of urban development. With this study the fact that aesthetics are associated with housing prices not only for occupiers, but also those of neighbors as an externality is proven. In case the association actually reflects a causal effect, policy is likely to account for influences of aesthetics beyond one's own home, outside the price system. The correct way of estimating the externalities, in order to not underprovide or overprovide for the effect or association, has to be enhanced and corrected to context. Policy makers could take the possible effects on nearby property owners, and the effect on their own imposed role as compensator, into account in their decision-making regarding considerations on whether to assign, keep or (partially) dispose of aesthetics committees.

This study falls or stands with the validity of the number of geotagged user-generated pictures on photo-sharing websites as a proxy for the aesthetic appeal of the urban environment. It might be that the number of photos within a 50-meter radius is not a good proxy for the aesthetic appeal of the urban environment as the number of photos rises above 30. Although Saiz, Salazar and Bernard (2018) carried out valid research to this phenomenon, confirming the proxy, the proxy might not be long-standing due to increased access to technology and social media. As can be seen in their study, the average amount of geotagged user-generated pictures per building on Flickr was 0.242 in 2011 and 0.612 in 2014 (Saiz, Salazar and Bernard, 2018). The average amount of images within 50 meters of a building for this study already rises to almost 9. For this reason, might the proxy be still valid, the association has to be reevaluated every certain time-period. On top of that, the number of pictures for every city or any part in the world might be substantially different, requiring a study such as these for every location, as this study can only be seen as valid for the cities of Lyon and Toulouse, or at most French landlocked cities. Further studies on the topic might develop a scale of attractiveness that is linked to the number of geotagged photos within a 50-meter radius, or perhaps quartiles or deciles of pictures at a certain location. The development of consistent measures that could be applied to different cities would be very useful. However, it appears that custom solutions are required. For example, cities at the seaside are found to have a large number of geotagged pictures at the waterfront, that could interfere with the validity of the proxy and ultimately the results.

A shortcoming for this study is that a large chunk of data had to be deleted due to the problem of values of whole transaction deals (e.g. apartment complexes) that were assigned to each individual building in those transactions. It could be assumed that most of these transactions were made by large investors, such as pension funds. With the removal of these observations, the predicted results delivered by the remaining observations might not be generalizable to groups of houses that transact in their entirety between large investors. Finally, this method of predicting the value that can be assigned to the aesthetic appeal of the urban environment by counting the number of geotagged images in the vicinity will presumably not be suitable for automated valuation models as the number of geotagged images can easily be manipulated by any person that is willing to drive up their real estate value. That there is a value associated with the aesthetic appeal of the urban environment, however, is proven by means of this study and can be taken into account by policy makers and scientists.

6. CONCLUSIONS

This paper has used the hedonic pricing method in order to estimate the association between the aesthetic appeal of the urban environment and nearby house prices in Lyon and Toulouse. Geolocation data of user-generated public images uploaded on online photo-sharing website Flickr were used to generate a proxy for the aesthetic appeal of the urban environment. The findings show that the aesthetic appeal of the urban environment does have a positive association with nearby house prices. However, the magnitude of this association is unclear as a result of that the magnitude of the association is possibly partially absorbed by the spatial control measures. Looking within a 50-meter radius around houses, findings show that having 1 to 10 pictures is associated with at least 3.95 percent higher house prices, having 11 to 20 pictures is associated with at least 4.67 percent higher house prices and having 21 to 30 pictures is associated with at least 5.25 percent higher house prices. In the urban environment, where houses are often built closely together, it may sum up to a very substantial value. This study contributes to the growing body of literature using publicly available user-generated data to examine phenomena. The findings of this study can inform policy makers who make decisions that involve the regulation of aesthetics in the urban environment, maximizing the well-being of urban citizens.

REFERENCES

Ahlfeldt, G. M. (2013) *Urbanity*.

Ahlfeldt, G. and Mastro, A. (2012) Valuing Iconic Design: Frank Lloyd Wright Architecture in Oak Park, Illinois, *Housing Studies*, 27(8), pp. 1079–1099.

Alonso, W. (1960) A Theory of the Urban Land Market, *Papers and Proceedings of the Regional Science Association*, 6, pp. 149–157.

Antoniou, V., Military, H., Service, G., Morley, J. G., Survey, O. and Haklay, M. (2010) Web 2.0 geotagged photos: assessing the spatial dimension of the phenomenon, *Geomatica*, 64(1), pp. 99–110.

Archined (2004) Honderd gemeenten deels welstandsvrij, [online] Available at: https://www.archined.nl/2004/07/honderd-gemeenten-deels-welstandsvrij/ (Accessed 5 October 2019).

Bae, C. C., Jun, M. and Park, H. (2003) The impact of Seoul's subway Line 5 on residential property values, *Transport Policy*, 10, pp. 85–94.

Benson, E. D., Hansen, J. L., Schwartz, A. L. and Smersh, G. T. (1998) Pricing Residential Amenities : The Value of a View, *Journal of Real Estate Finance and Economics*, 16(1), pp. 55–73.

Besanko, D., Dranove, D., Shanley, M. and Schaefer, S. (2013) *Economics of Strategy*, 6th ed, Singapore, John Wiley & Sons.

Bond, M. T., Seiler, V. L. and Seiler, M. J. (2002) Residential Real Estate Prices: A Room with a View, *Journal of Real Estate Research*, 23(1), pp. 129–138.

Bourassa, S. C., Hoesli, M. and Sun, J. (2005) The Price of Aesthetic Externalities, *Journal of Real Estate Literature*, 13(2), pp. 14-167–190.

Bourassa, S. C., Hoesli, M. and Sun, J. (2004) What's in a view?, *Environment and Planning A*, 36(8), pp. 1427–1450.

BouwWereld (2011) Blog documenteert lelijkheid Belgische huizen, [online] Available at: https://www.bouwwereld.nl/bouwkennis/blog-documenteert-stuitende-lelijkheid-belgische-huizen/ (Accessed 5 October 2019).

Brooks, C. and Tsolacos, S. (2010) *Real Estate Modelling and Forecasting*, Cambridge, Cambridge University Press.

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

Caudenys, H. (2015) *Ugly Belgian Houses*, 1st ed, Borgerhoff & Lamberigts.

Celik, M. H. and Yankaya, U. (2006) The impact of rail transit investment on the residential property values in developing countries: The case of Izmir Subway, Turkey, *Property Management*, 24(4), pp. 369–382.

Cheshire, B. P. and Sheppard, S. (1995) On the Price of Land and the Value of Amenities, *Economica*, 62(246), pp. 247–267.

Daams, M. N., Sijtsma, F. J. and van der Vlist, A. J. (2016) The Effect of Natural Space on Nearby Property Prices : Accounting for Perceived Attractiveness, *Land Economics*, 92(3), pp. 389–410.

Darling, A. H. (1973) Measuring Benefits Generated by Urban Water Parks, *Land Economics*, 49(1), pp. 22–34.

DiPasquale, D. and Wheaton, C. (1996) *Urban Economics and Real Estate Markets*, New Jersey, Pearson.

Droit-Finances (2019) Construire avec ou sans architecte, [online] Available at: https://droit-finances.commentcamarche.com/contents/667-construire-avec-ou-sans-architecte#aspect-exterieur (Accessed 11 October 2019).

Etalab (2019) Etalab, [online] Available at: data.gouv.fr/en/organizations/etalab.

Federale Overheidsdienst Economie (2019a) Bouwen Met Een Architect, [online] Available at: https://economie.fgov.be/nl/themas/consumentenbescherming/bouwen/bouwen-met-een-architect (Accessed 11 October 2019).

Federale Overheidsdienst Economie (2019b) De Architecten, [online] Available at: https://economie.fgov.be/nl/over-de-fod/missie-visie-waarden/waarden (Accessed 11 October 2019).

Federatie Ruimtelijke Kwaliteit (2016) (Bijna) alles over welstand en ruimtelijke kwaliteit, [online] Available at: https://www.ruimtelijkekwaliteit.nl/federatie_ruimtelijke_kwaliteit/over_welstand (Accessed 4 October 2019).

Gemeente Gouda (2004) *Welstandsvrij bouwen in Gouda*, Gouda, [online] Available at: https://www.gouda.nl/dsresource?objectid=1283594c-6dca-45b4-ab88-d8ac6f0de8d1.

Gillen, K., Thibodeau, T. and Wachter, S. (2001) Anisotropic Autocorrelation in House Prices, *Journal of Real Estate Finance and Economics*, 23(1), pp. 5–30.

Glaeser, E. L., Kolko, J. and Saiz, A. (2001) Consumer City, *Journal of Economic Geography*, 1, pp. 27–50.

Halvordsen, R. and Palmquist, R. (1980) The Interpretation of Dummy Variables in Semi-logarithmic Equations, *American Economic Review*, 70(3), pp. 474–475.

Han, H. (2014) The Impact of Abandoned Properties on Nearby Property Values, *Housing Policy Debate*, Taylor & Francis, 24(2), pp. 311–334, [online] Available at: http://dx.doi.org/10.1080/10511482.2013.832350.

JUSTEL (2019) Geconsolideerde Wetgeving: Wet op de bescherming van den titel en het beroep architect, [online] Available at: http://www.ejustice.just.fgov.be/cgi_loi/loi_a.pl?language=nl&caller=list&cn=1939022030&la=n&fromtab=wet&sql=dt=%27wet%27&tri=dd+as+rank&rech=1&numero=1 (Accessed 11 October 2019).

Lazrak, F., Nijkamp, P., Rietveld, P. and Rouwendal, J. (2014) The market value of cultural heritage in urban areas: An application of spatial hedonic pricing, *Journal of Geographical Systems*, 16(1), pp. 89–114.

Levi, D. J. (2016) Does History Matter? Perceptions and Attitudes Toward Fake Historic Architecture and Historic Preservation, *Journaal of Architectural and Planning Research*, 22(2), pp. 148–159.

Millhouse, J. A. (2005) Assessing the Effect of Architectural Design on Real Estate Values: A Qualitative Approach,.

Ming, Y. S. and Hian, C. C. (2005) Obstruction of View and its Impact on Residential Apartment Prices, *Pacific Rim Property Research Journal*, 11(3), pp. 299–315.

Montero, J. M., Fernández-Avilés, G. and Mínguez, R. (2018) Estimating environment impacts on housing prices, *Environmetrics*, 29(5–6), pp. 1–12.

Nicholls, S. and Crompton, J. L. (2018) The contribution of scenic views of, and proximity to, lakes and reservoirs to property values, *Lakes & Reservoirs: Research & Management*, 23(1), pp. 63–78.

Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C. and González, M. C. (2015) Urban magnetism through the lens of geo-tagged photography, *EPJ Data Science*, 4(5), [online] Available at: http://dx.doi.org/10.1140/epjds/s13688-015-0043-3.

Palmquist, R. B. (2005) Property Value Models, In Göran-Mähler, K. and Vicent, J. R. (eds.), *Handbook of Environmental Economics*, 2nd ed, Amsterdam: North Holland, pp. 763–819.

Plattner, R. H. and Campbell, T. J. (1978) A Study of the Effect of Water View on Site Value, *Appraisal Journal*, 46(1), pp. 20–25.

Rodriguez, M. and Sirmans, C. F. (1994) Quantifying the Value of a View in Singe-Family Housing Markets, *Appraisal Journal*, 62(1), pp. 600–603.

Rosen, S. (1974) Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, 82(1), pp. 34–55.

Ruijgrok, E. C. M. (2006) The three economic values of cultural heritage: a case study in the Netherlands, *Journal of Cultural Heritage*, 7(3), pp. 206–213.

Saiz, A., Salazar, A. and Bernard, J. (2018) Crowdsourcing architectural beauty: Online photo frequency predicts building aesthetic ratings, *PLOS ONE*, pp. 1–15.

Seresinhe, C. I., Moat, H. S. and Preis, T. (2018) Quantifying scenic areas using crowdsourced data, *Urban Analytics and City Science*, 45(3), pp. 567–582.

Song, Y. and Knaap, G. (2003) New urbanism and housing values: a disaggregate assessment, *Journal of Urban Economics*, 54, pp. 218–238.

Tiwari, P. and White, M. (2010) *International Real Estate Economics*, New York, Palgrave.

Trouw (2019) De commissie die bepaalt welke kleur een huis wordt, blijkt vooral een smaakpolitie, *Trouw*, [online] Available at: https://www.trouw.nl/binnenland/de-commissie-die-bepaalt-welke-kleur-een-huis-wordt-blijkt-vooral-een-smaakpolitie~b71bc77b/ (Accessed 5 October 2019).

Trouw (2011) Het interesseert sommigen geen moer wat ze neerzetten, *Trouw*, [online] Available at: https://www.trouw.nl/nieuws/het-interesseert-sommigen-geen-moer-wat-ze-neerzetten~b046a16d/ (Accessed 5 October 2019).

Upton, G. and Cook, I. (2008) *A Dictionary of Statistics*, Oxford, OUP.

Wilkinson, R. K. (1973) House Prices and the Measurement of Externalities, *The Economic Journal*, 83(329), pp. 72–86.

ZIN (2019) Beeldregie versus welstandsvrij, [online] Available at: https://www.zelfbouwinnederland.nl/informatie/project-voorbereiden/beeldkwaliteitseisen/beeldregie-versus-welstandsvrij (Accessed 5 October 2019).

APPENDIX A: Transforming variables into natural logarithms

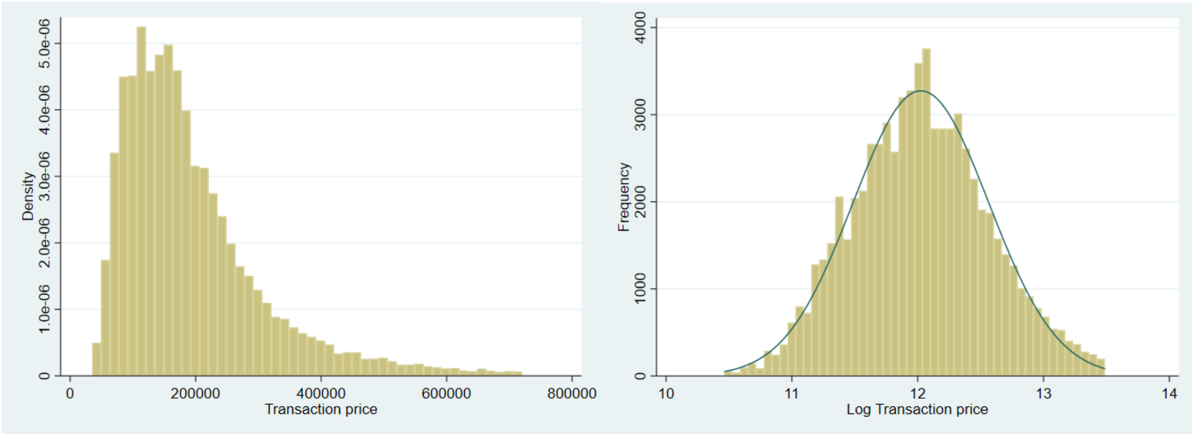

Figure 1: Histograms of transaction price before and after transforming

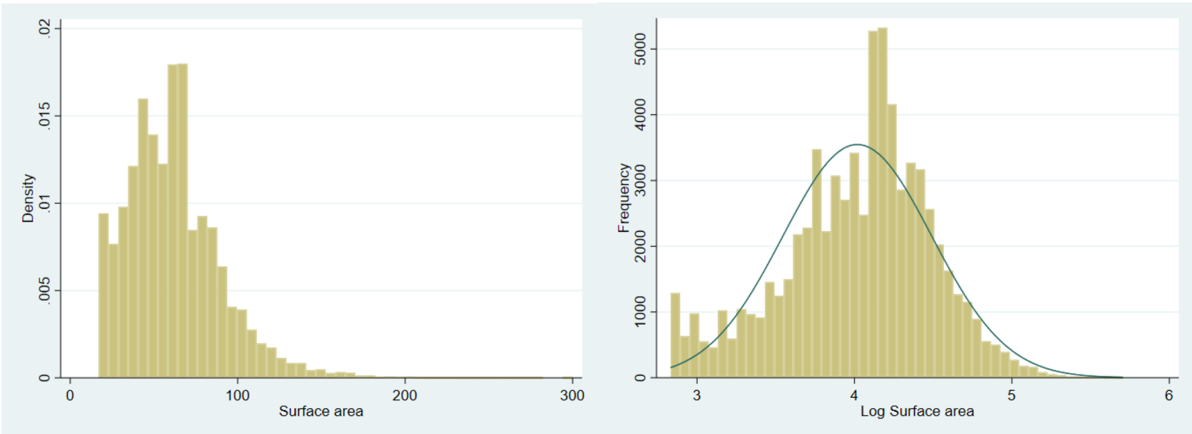

Figure 2: Histograms of surface area before and after transforming

APPENDIX B: Testing OLS assumptions

The use of an OLS estimation method requires the following assumptions (Brooks and Tsolacos, 2010):

Table 3: OLS assumptions

Assumption	Description
1. $E(\varepsilon t) = 0$ Linearity	The average value of the errors is zero
2. $Var(\varepsilon t) = \sigma < \infty$ Homoscedasticity	The variance of the errors is constant and finite
3. $Cov (\varepsilon i, \varepsilon j) = 0$ for $i = j$ No autocorrelation	The errors are statistically independent of one another
4. $Cov (\varepsilon t, xt) = 0$ Independence	There is no relationship between the errors and the corresponding x-variables
5. εt $N(0, \sigma 2)$ Normality	The errors are approximately normally distributed

If these assumptions are not met, OLS may run into problems (Brooks and Tsolacos, 2010). Brooks and Tsolacos (2010) argue that when assumptions 1 to 4 hold, the estimated coefficients determined by OLS are BLUE, which stands for Best Linear Unbiased Estimator. The estimated coefficients will be consistent, unbiased and efficient, meaning they approximately equal their true value (Brooks and Tsolacos, 2010). With BLUE estimated coefficients, conclusions can be drawn about relationships between dependent and independent variables. The dataset is tested for the OLS assumptions. It appeared that assumption 2, homoscedasticity, was violated as the variance of the errors was not constant. This is fixed by running the regression with robust standard errors. Further, assumption 3 was violated as autocorrelation was detected. By clustering the errors at the postal code level, the errors are made statistically independent. The remaining assumptions were met without any interventions. See below for further discussion.

Assumption 1: Linearity

The linearity assumption will never be violated if a constant term is included in the regression equation (Brooks and Tsolacos, 2010). Stata, a statistical software package that is used for this study, automatically includes a constant term in all regressions, resulting in the fulfilling of this assumption.

Assumption 2: Homoscedasticity

Rvfplot is a visual test for homoscedasticity. The result shows that the errors might be heteroscedastic. The Bruesch-Pagan/Cook-Weisberg test for heteroscedasticity is carried out as well. The null hypothesis is that the variance of the errors is constant (and thus homoscedastic). The null hypothesis is rejected as the p-value is below 0.05 (0.0000). This means that the variance of the errors is not constant. The solution to meet this assumption anyways is to run the regression with robust standard errors, which is implemented in the research design.

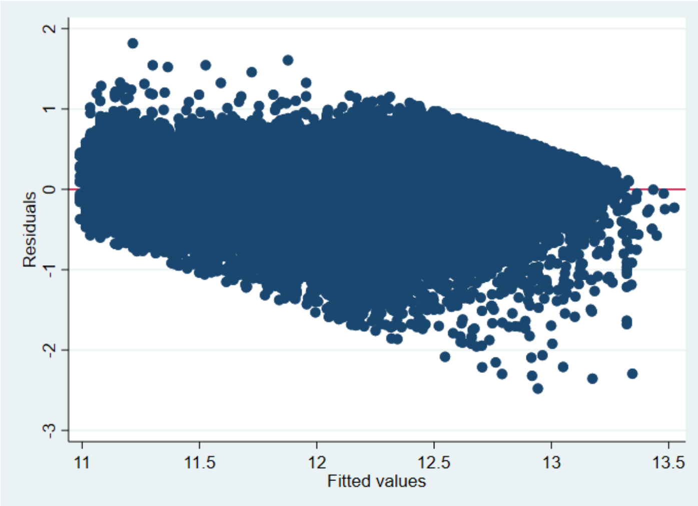

Figure 3: Rvf plot (test for homoscedasticity)

Assumption 3: No autocorrelation

Spatial autocorrelation is very common in real estate research datasets, as houses close to each other often have similar characteristics (Gillen, Thibodeau and Wachter, 2001). Therefore, it is generally accepted to assume spatial autocorrelation in real estate data (Gillen, Thibodeau and Wachter, 2001). A solution for autocorrelation is running the regression with clustered standard errors. This solution is adapted for this study: the standard errors are clustered at the postal code level in two models.

Assumption 4: Independence

Endogeneity is tested by the Durbin-Wu-Hausman test. In this test, the null-hypothesis is that there is no dependence between the errors and corresponding x-variables. The tests for the different dependent x-variables all result in non-significant coefficients (see table 4), meaning the null-hypothesis cannot be rejected. There are no relationships between the errors and the corresponding x-variables (no endogeneity).

Table 4: Durbin-Wu-Hausman test results

	(1) Log Transaction price	(2) Log Transaction price	(3) Log Transaction price	(4) Log Transaction price
Errors 1-10 photos	5.536 (4.671)			
Errors 11-20 photos		124.8 (460.4)		
Errors 21-30 photos			41.03 (34.62)	
Errors 31+ photos				15.64 (13.19)
Constant	10.84*** (1.586)	11.94 (12.94)	9.020*** (0.0902)	9.868*** (0.771)
Observations	70,022	70,022	70,022	70,022
R-squared	0.654	0.654	0.654	0.654

In parentheses are the standard errors that are robust to heteroscedasticity and clustered at the postal code level.
*, **, *** Significance at 10%, 5% and 1% respectively. Only the variables of interest and the constant are showed.

Assumption 5: Normality

Normality is tested by the Kernel density plot and a standardized normal probability plot (pnorm). The Kernel density plot needs to approximately overlay the normal density. This is approximately the case, as can be seen in figure 4. Further, the standardized normal probability plot (pnorm) needs to approximately follow the line in order to be normally distributed. This is also approximately the case (see figure 5), so the errors are normally distributed.

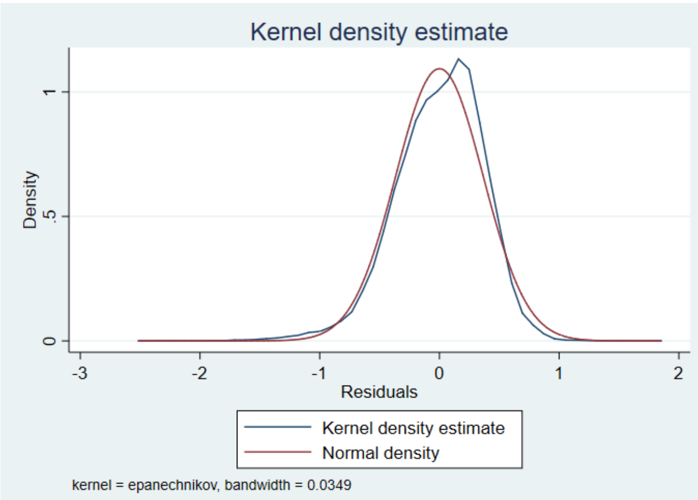

Figure 4: Kernel density plot

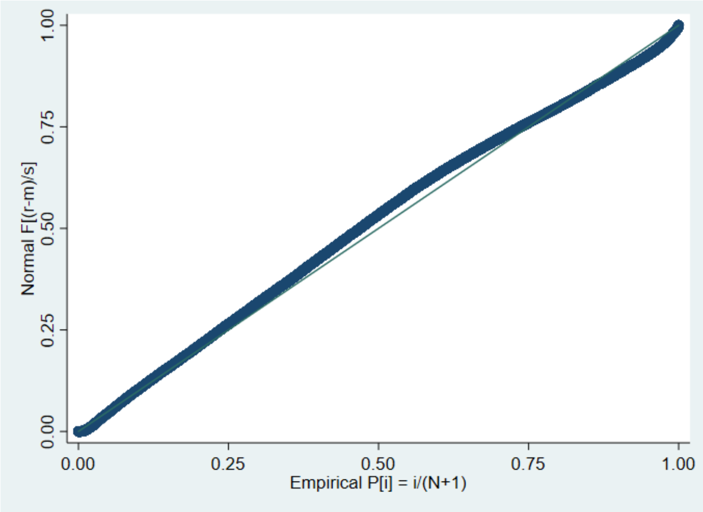

Figure 5: Standardized normal probability plot (pnorm)

APPENDIX C: Assessing multicollinearity

Table 5: Correlation matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Transaction price	1.000											
(2) Transaction year	0.020	1.000										
(3) Surface area	0.757*	0.016	1.000									
(4) Rooms	0.597*	0.034	0.860*	1.000								
(5) Building type	-0.308	-0.044	-0.340	-0.333	1.000							
(6) Postal code	0.220	-0.157	0.094	-0.000	0.182	1.000						
(7) Number of photos <50m	0.071	-0.017	0.005	-0.044	0.075	0.124	1.000					
(8) D Photos 0	-0.102	0.044	0.044	0.126	-0.158	-0.239	-0.374	1.000				
(9) D Photos 1-10	0.046	-0.036	-0.050	-0.091	0.116	0.168	-0.095	-0.792	1.000			
(10) D Photos 11-20	0.039	-0.008	-0.010	-0.046	0.035	0.055	0.077	-0.224	-0.114	1.000		
(11) D Photos 21-30	0.045	-0.007	0.009	-0.023	0.035	0.053	0.127	-0.168	-0.085	-0.024	1.000	
(12) D Photos 31+	0.072	-0.012	0.009	-0.041	0.062	0.105	0.858	-0.301	-0.153	-0.043	-0.032	1.000

An important consideration is whether there is multicollinearity in the data. This is important in order to perform a multiple linear regression. Multicollinearity is present when independent variables are highly correlated (Brooks and Tsolacos, 2010). Multicollinearity will make the regression very sensitive to small changes in the specification and makes the confidence intervals for the parameters very wide, leading to inappropriate conclusions from significance tests (Brooks and Tsolacos, 2010). In the correlation matrix above, high correlations are marked with an asterisk. High correlations with the dependent variable (transaction price) can be ignored, as they do not cause multicollinearity. It can be seen that there is a relatively high correlation between the surface area and the number of rooms. However, these variables are both important and therefore not omitted, as omitted variables result in omitted variable bias (Brooks and Tsolacos, 2010). On top of that, the correlation is not high enough to cause multicollinearity. All VIF values are well below 10, the standard benchmark (Brooks and Tsolacos, 2010), indicating there is no multicollinearity in the data.

APPENDIX D: PHOTO FREQUENCY DISTRIBUTION MAPS

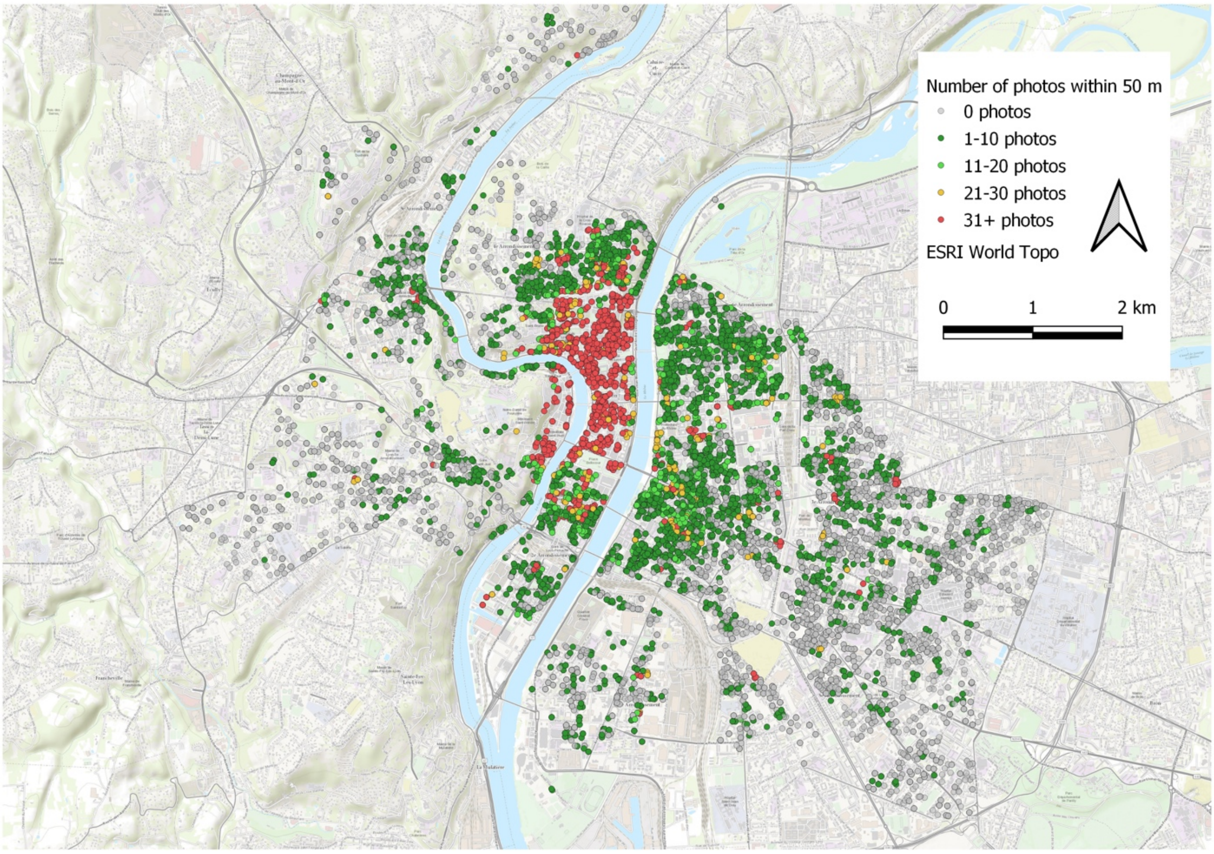

Figure 6: Spatial distribution of the number of photos within 50 meters of houses in Lyon, France

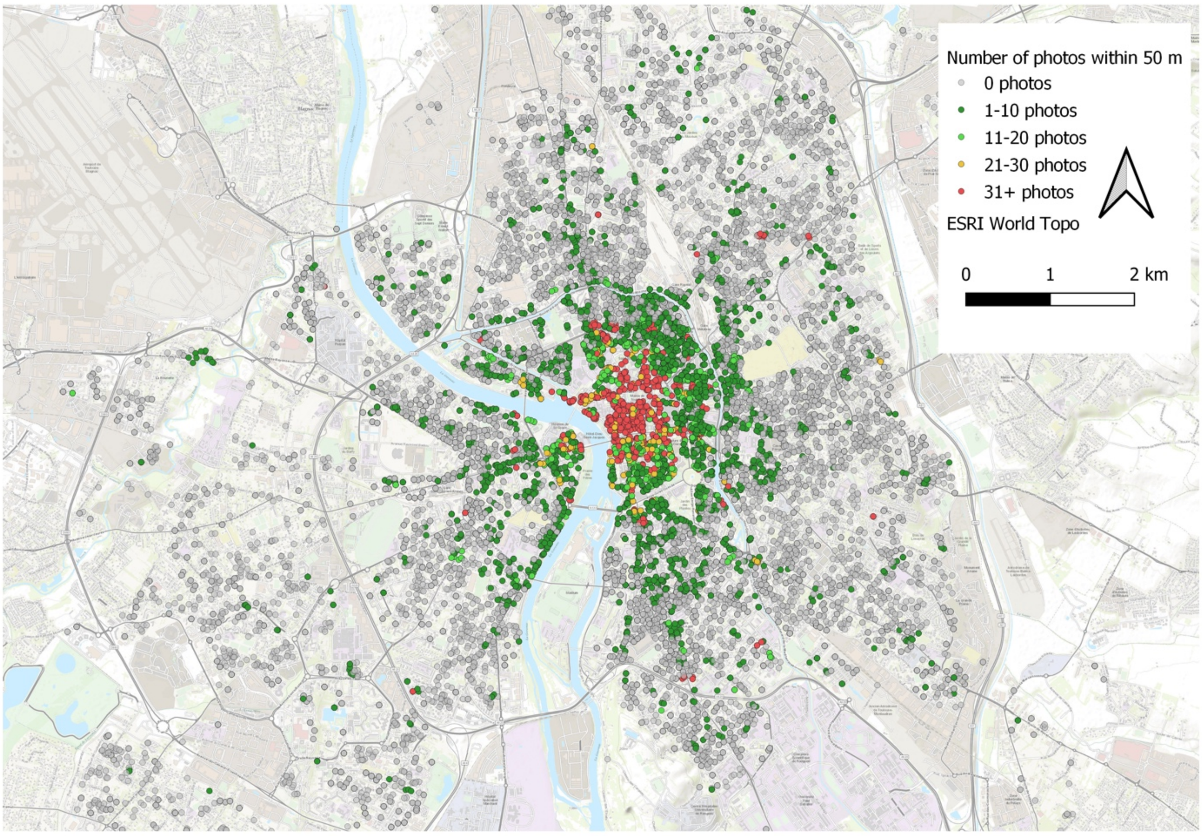

Figure 7: Spatial distribution of the number of photos within 50 meters of houses in Toulouse, France

APPENDIX E: STATA SYNTAX

```
*MSc Thesis JW van der Kam

clear all

*Pathway to data
cd "X:\My Documents\Scriptiedata\Combined data"
use CompleteData

*Set place to save results
cd "X:\My Documents\Scriptiedata\Results"

*Install outreg2
sysdir set PLUS "X:\My Documents\Scriptiedata\Results"
ssc install outreg2, replace

*Install asdoc
ssc install asdoc, replace
net install asdoc, from(http://fintechprofessor.com) replace


*Preparing the dataset

*Drop variables that are not useful
drop fid
drop ancien_code_commune ancien_nom_commune
drop lot1_numero lot1_surface_carrez lot2_numero lot2_surface_carrez
lot3_numero lot3_surface_carrez lot4_numero lot4_surface_carrez lot5_numero
lot5_surface_carrez nombre_lots
drop code_nature_culture_speciale nature_culture_speciale
code_nature_culture nature_culture
drop layer path
drop ancien_id_parcelle
drop numero_disposition
drop numero_volume
drop surface_terrain

*Renaming variables
rename valeur_fonciere transaction_price
rename adresse_numero address_number
rename adresse_suffixe address_suffix
rename adresse_nom_voie address_name
rename adresse_code_voie address_code
rename code_postal postal_code
rename code_commune commune_code
rename nom_commune commune_name
rename code_departement department_code
rename id_parcelle id_parcel
rename surface_reelle_bati real_built_surface
rename nombre_pieces_principales main_rooms
rename numberphotos number_photos

*Drop outbuildings
drop if code_type_local == 3

*Drop duplicates
duplicates report id_mutation
duplicates tag id_mutation, generate(duplicatestag)
keep if duplicatestag == 0
*Drop industrial and commercial observations
drop if code_type_local == 4
```

```
drop if code_type_local == .

*Remove outliers transaction prices
sum transaction_price, detail
keep if inrange(transaction_price, r(p1), r(p99))
sum transaction_price, detail

*Drop observations without coordinates
drop if latitude == .
drop if longitude == .

*Drop observations without transaction price information
drop if transaction_price == .

*Drop observations without year
drop if year == .

*Drop observations without postal code
drop if postal_code == .

*Drop observations without any room information
drop if main_rooms == .
drop if main_rooms == 0

*Drop observations without any surface information or unreasonable surface
sum real_built_surface, detail
drop if real_built_surface == .
drop if real_built_surface < 17

*Remove outliers rooms
sum main_rooms, detail
drop if main_rooms > 6
sum main_rooms, detail

*Generate photo classes
gen photos_0 = 0
gen photos_1_10 = 0
gen photos_11_20 = 0
gen photos_21_30 = 0
gen photos_31_plus = 0

replace photos_0 = 1 if number_photos == 0

replace photos_1_10 = 1 if number_photos == 1
replace photos_1_10 = 1 if number_photos == 2
replace photos_1_10 = 1 if number_photos == 3
replace photos_1_10 = 1 if number_photos == 4
replace photos_1_10 = 1 if number_photos == 5
replace photos_1_10 = 1 if number_photos == 6
replace photos_1_10 = 1 if number_photos == 7
replace photos_1_10 = 1 if number_photos == 8
replace photos_1_10 = 1 if number_photos == 9
replace photos_1_10 = 1 if number_photos == 10

replace photos_11_20 = 1 if number_photos == 11
replace photos_11_20 = 1 if number_photos == 12
replace photos_11_20 = 1 if number_photos == 13
replace photos_11_20 = 1 if number_photos == 14
replace photos_11_20 = 1 if number_photos == 15
replace photos_11_20 = 1 if number_photos == 16
replace photos_11_20 = 1 if number_photos == 17
replace photos_11_20 = 1 if number_photos == 18
replace photos_11_20 = 1 if number_photos == 19
replace photos_11_20 = 1 if number_photos == 20
```

```
replace photos_21_30 = 1 if number_photos == 21
replace photos_21_30 = 1 if number_photos == 22
replace photos_21_30 = 1 if number_photos == 23
replace photos_21_30 = 1 if number_photos == 24
replace photos_21_30 = 1 if number_photos == 25
replace photos_21_30 = 1 if number_photos == 26
replace photos_21_30 = 1 if number_photos == 27
replace photos_21_30 = 1 if number_photos == 28
replace photos_21_30 = 1 if number_photos == 29
replace photos_21_30 = 1 if number_photos == 30

replace photos_31_plus = 1 if number_photos >= 31

*Increase matsize
set matsize 11000

*Labels for exporting tables and figures
label variable transaction_price "Transaction price"
label variable real_built_surface "Surface area"
label variable main_rooms "Rooms"
label variable code_type_local "Building type"
label variable postal_code "Postal code"
label variable address_code_city "Address code"
label variable number_photos "Number of photos <50m"
label variable photos_0 "Photos 0"
label variable photos_1_10 "D Photos 1-10"
label variable photos_11_20 "D Photos 11-20"
label variable photos_21_30 "D Photos 21-30"
label variable photos_31_plus "D Photos 31+"
label variable year "Transaction year"
label variable log_price "Log Transaction price"
label variable log_surface "Log Surface area"

*Histograms
hist transaction_price
graph export transaction_price.png, replace
hist real_built_surface
graph export real_built_surface.png, replace

*Generate natural logarhitms
gen log_price = log(transaction_price)
gen log_surface = log(real_built_surface)

*Histograms logs
hist log_price, frequency normal
graph export log_price.png, replace
hist log_surface, frequency normal
graph export log_surface.png, replace


*Descriptive statistics

gen photos_0_ds = number_photos if photos_0 == 1
label variable photos_0_ds "Photos 0"


gen photos_1_10_ds = number_photos if photos_1_10 == 1
label variable photos_1_10_ds "Photos 1-10"

gen photos_11_20_ds = number_photos if photos_11_20 == 1
label variable photos_11_20_ds "Photos 11-20"

gen photos_21_30_ds = number_photos if photos_21_30 == 1
```

```
label variable photos_21_30_ds "Photos 21-30"

gen photos_31_plus_ds = number_photos if photos_31_plus == 1
label variable photos_31_plus_ds "Photos 31+"

sum transaction_price year real_built_surface main_rooms code_type_local
number_photos photos_0_ds photos_1_10_ds photos_11_20_ds photos_21_30_ds
photos_31_plus_ds photos_0 photos_1_10 photos_11_20 photos_21_30
photos_31_plus

*Export descriptive statistics to document
asdoc sum transaction_price year real_built_surface main_rooms
code_type_local number_photos photos_0_ds photos_1_10_ds photos_11_20_ds
photos_21_30_ds photos_31_plus_ds photos_0 photos_1_10 photos_11_20
photos_21_30 photos_31_plus, abb(.) label

*Correlation matrix
corr transaction_price year real_built_surface main_rooms code_type_local
postal_code number_photos photos_0 photos_1_10 photos_11_20 photos_21_30
photos_31_plus

*Export correlation matrix to document
asdoc cor transaction_price year real_built_surface main_rooms
code_type_local postal_code number_photos photos_0 photos_1_10 photos_11_20
photos_21_30 photos_31_plus, replace abb(.) label

*VIF values
reg log_price log_photos photos_1_10 photos_11_20 photos_21_30
photos_31_plus log_surface i.main_rooms i.code_type_local i.year
postal_code_1000
estat vif


*Analysis

*Models 1, 2, 3, 4, 5
areg log_price log_surface i.main_rooms i.code_type_local i.year,
absorb(postal_code) robust cluster(postal_code)
areg log_price number_photos log_surface i.main_rooms i.code_type_local
i.year, absorb(postal_code) robust cluster(postal_code)
areg log_price photos_1_10 photos_11_20 photos_21_30 photos_31_plus
log_surface i.main_rooms i.code_type_local i.year, absorb(postal_code)
robust cluster(postal_code)
reg log_price number_photos log_surface i.main_rooms i.code_type_local
i.year, robust
reg log_price photos_1_10 photos_11_20 photos_21_30 photos_31_plus
log_surface i.main_rooms i.code_type_local i.year, robust

*Exporting models to a document
areg log_price log_surface i.main_rooms i.code_type_local i.year,
absorb(postal_code) robust cluster(postal_code)
outreg2 using outputregressions2.doc, label replace
areg log_price number_photos log_surface i.main_rooms i.code_type_local
i.year, absorb(postal_code) robust cluster(postal_code)
outreg2 using outputregressions2.doc, label
areg log_price photos_1_10 photos_11_20 photos_21_30 photos_31_plus
log_surface i.main_rooms i.code_type_local i.year, absorb(postal_code)
robust cluster(postal_code)
outreg2 using outputregressions2.doc, label
reg log_price number_photos log_surface i.main_rooms i.code_type_local
i.year, robust
outreg2 using outputregressions2.doc, label
reg log_price photos_1_10 photos_11_20 photos_21_30 photos_31_plus
log_surface i.main_rooms i.code_type_local i.year, robust
```

```
outreg2 using outputregressions2.doc, label


*Testing OLS assumptions

*Assumption 2
rvfplot, yline(0)
graph export rvfplot.png, replace
estat hettest

*Assumption 4
areg log_price photos_1_10 photos_11_20 photos_21_30 photos_31_plus
log_surface i.main_rooms i.code_type_local i.year, absorb(postal_code)
robust cluster(postal_code)
areg photos_1_10 photos_11_20 photos_21_30 photos_31_plus log_surface
i.main_rooms i.code_type_local i.year, absorb(postal_code) robust
cluster(postal_code)
predict errors_1_10, resid
label variable errors_1_10 "Errors 1-10 photos"
areg log_price errors_1_10 photos_1_10 photos_11_20 photos_21_30
photos_31_plus log_surface i.main_rooms i.code_type_local i.year,
absorb(postal_code) robust cluster(postal_code)
outreg2 using assumption4x4.doc, label

areg log_price photos_1_10 photos_11_20 photos_21_30 photos_31_plus
log_surface i.main_rooms i.code_type_local i.year, absorb(postal_code)
robust cluster(postal_code)
areg photos_11_20 photos_1_10 photos_21_30 photos_31_plus log_surface
i.main_rooms i.code_type_local i.year, absorb(postal_code) robust
cluster(postal_code)
predict errors_11_20, resid
label variable errors_11_20 "Errors 11-20 photos"
areg log_price errors_11_20 photos_1_10 photos_11_20 photos_21_30
photos_31_plus log_surface i.main_rooms i.code_type_local i.year,
absorb(postal_code) robust cluster(postal_code)
outreg2 using assumption4x4.doc, label

areg log_price photos_1_10 photos_11_20 photos_21_30 photos_31_plus
log_surface i.main_rooms i.code_type_local i.year, absorb(postal_code)
robust cluster(postal_code)
areg photos_21_30 photos_1_10 photos_11_20 photos_31_plus log_surface
i.main_rooms i.code_type_local i.year, absorb(postal_code) robust
cluster(postal_code)
predict errors_21_30, resid
label variable errors_21_30 "Errors 21-30 photos"
areg log_price errors_21_30 photos_1_10 photos_11_20 photos_21_30
photos_31_plus log_surface i.main_rooms i.code_type_local i.year,
absorb(postal_code) robust cluster(postal_code)
outreg2 using assumption4x4.doc, label

areg log_price photos_1_10 photos_11_20 photos_21_30 photos_31_plus
log_surface i.main_rooms i.code_type_local i.year, absorb(postal_code)
robust cluster(postal_code)
areg photos_31_plus photos_1_10 photos_11_20 photos_21_30 log_surface
i.main_rooms i.code_type_local i.year, absorb(postal_code) robust
cluster(postal_code)
predict errors_31_plus, resid
label variable errors_31_plus "Errors 31+ photos"
areg log_price errors_31_plus photos_1_10 photos_11_20 photos_21_30
photos_31_plus log_surface i.main_rooms i.code_type_local i.year,
absorb(postal_code) robust cluster(postal_code)
outreg2 using assumption4x4.doc, label
*Assumption 5
predict r, resid
```

```
hist r, normal
kdensity r, normal
pnorm r
qnorm r


*Chow test

areg log_price number_photos log_surface i.main_rooms i.code_type_local
i.year, absorb(postal_code) robust cluster(postal_code)

areg log_price number_photos log_surface i.main_rooms i.code_type_local
i.year if department_code == 31, absorb(postal_code) robust
cluster(postal_code)
areg log_price number_photos log_surface i.main_rooms i.code_type_local
i.year if department_code == 69, absorb(postal_code) robust
cluster(postal_code)

gen d = 0
replace d = 1 if department_code == 69
gen photos_d = number_photos * d
areg log_price number_photos photos_d d log_surface i.main_rooms
i.code_type_local i.year, absorb(postal_code) robust cluster(postal_code)

test _b[photos_d]=0, notest
test _b[d]=0, accum
```