Modelling Spatial Scale and Heterogeneity in Rotterdam Housing Market Using Multiscale Geographically Weighted Regression

Authors: Naftali Feddes
Submitted: 30. June 2023
Published: 17. July 2023
Volume: 10
Issue: 4
Affiliation: Erasmus School of Economics, Rotterdam, Netherlands
Languages: English
Keywords: Rotterdam Housing Market, Spatial Heterogeneity, Local and Global Effects, Residential Housing Prices
Categories: Demetrios Project, News and Views, Visual Arts, Architecture and Design
DOI: 10.17160/josha.10.4.915

Abstract:

This study investigates the spatial heterogeneity of the Rotterdam housing market, as well as the spatial scale at which the processes that explain housing prices operate. I use data on 3632 residential properties that were sold in Rotterdam in 2018 provided by the NVM. Using a recently (2017) developed technique called Multiscale Geographically Weighted Regression (MGWR), I show that certain factors explaining residential housing prices operate on different spatial scales, indicating the existence of relatively local and more global effects. Lot size and the state of maintenance of the house exterior seem to operate on a local scale, whereas the degree of isolation and number of rooms appear to be global. Comparison of the MGWR model with a regular GWR with fixed bandwidths and a global hedonic model with location fixed effects, show that the MGWR best fits the data, i.e. a model that allows the bandwidths to vary for each parameter outperforms the models that do not allow this.

JOSHA
josha.org

Journal of Science, Humanities and Arts

JOSHA is a service that helps scholars, researchers, and students discover, use, and build upon a wide range of content

Master Thesis - ESE

Modelling Spatial Scale and Heterogeneity in the Rotterdam Housing Market using Multiscale Geographically Weighted Regression

Author: Naftali Feddes *Supervisor*: Nicola Cortinovis

April 21, 2020

Abstract

This study investigates the spatial heterogeneity of the Rotterdam housing market, as well as the spatial scale at which the processes that explain housing prices operate. I use data on 3632 residential properties that were sold in Rotterdam in 2018 provided by the NVM. Using a recently (2017) developed technique called Multiscale Geographically Weighted Regression (MGWR), I show that certain factors explaining residential housing prices operate on different spatial scales, indicating the existence of relatively local and more global effects. Lot size and the state of maintenance of the house exterior seem to operate on a local scale, whereas the degree of isolation and number of rooms appear to be global. Comparison of the MGWR model with a regular GWR with fixed bandwidths and a global hedonic model with location fixed effects, show that the MGWR best fits the data, i.e. a model that allows the bandwidths to vary for each parameter outperforms the models that do not allow this.

The realization of this thesis has been made easier by a number of people, and I'd therefore like to extend my gratitude to the following individuals: Nicola Cortinovis, for introducing me to spatial economics and for his considerate supervision and feedback; Jeroen van Haaren, for providing me with the data and for additional insights in the Rotterdam housing market; Taylor Oshan, for his time and helpful comments on circumventing technical constraints in the modelling environment; last but not least, to my friends Marc Boef and Scott Reuvers for sparring with me on programming matters and for their loyal company in the university library.

Contents

1 Introduction

Real estate valuation dates back to at least the Roman civilization (Matthews, 1921), and unlike some of the objects under valuation, the appraisal criteria seem to have withstood the test of time. The ancient Greeks already established the importance of size and location, as larger houses closer to the agora appeared to yield a higher return than their smaller and more remote counterparts (Nevett, 2000). Although it is unclear what methods they employed, a property's location today is as relevant to its price as it was then. However, despite the advancements of computational software for real estate appraisal (McCluskey, McCord, Davis, Haran & McIlhatton, 2013), location is not always properly modelled. In the often used sales comparison approach (Kettani & Oral, 2015), where the value of the subject property is assumed to bear close resemblance to the selling prices of similar properties in its vicinity, location is not always quantified because it is assumed that experts possess sufficient knowledge on the location (Wyatt, 1997). These experts' judgment may be subjective and less than perfect (Ramsland & Markham, 1998; Yeh & Hsu, 2018). A more quantitative and objective approach is the hedonic pricing model, which lacks the human bias and is more suitable for (mass) appraisal, or as a benchmark for other models (Yeh & Hsu, 2018). In hedonic applications the house price is assumed to be a function of the implicit value of the houses' characteristics, sometimes further divided in categories like structural, accessibility, or neighbourhood characteristics (Fik et al. 2003). Even though hedonic regression models are better able to incorporate the locational (or spatial) effect, oftentimes through regional fixed effects dummies, they are not able to perfectly capture the spatial heterogeneity that is inherent in housing markets (Helbich, Brunauer, Vaz, & Nijkamp, 2014). This is because the strong spatial patterns in the supply of specific housing characteristics in metropolitan areas do not always obey the boundaries that location dummies impose.

To better capture the spatial patterns that emerge, we can resort to non-parametric models such as geographically weighted regression (GWR), where the nature of the parameters is dependent on the data and not fixed beforehand. However, even in GWR the same spatial scale is used for all the parameters (e.g. McMillen & Redfearn, 2010), while it is very well possible that some factors explaining house prices operate on a different spatial scale than others. For instance, the presence of gardens in nearby houses may provide positive spillovers for the object under sale, while gardens that are further away from the object under sale are less likely to affect the price. There still is a scarcity in the literature with regards to the geographic scale over which different economic processes operate, including those processes that govern real estate prices (Fotheringham, Yang, & Kang, 2017). In fact, Helbich et al. (2014) state that there is no empirical consensus as to which factors can be considered local (i.e. operating on a relatively small scale) and which global.

The aim of this study is to investigate these scales, i.e. the spatial scales at which processes operate in the determination of housing prices. Researching this is important because GWR results in biased estimates and potentially not justified conclusions if applied to data with different scales (Wei & Qi, 2012). Furthermore, it provides insights in processes that throughout the literature are nearly always assumed to be on the same scale. To demonstrate the spatial heterogeneity a recently developed method called Multiscale Geographically Weighted Regression (MGWR) is employed. Regular GWR uses a locational attribute in observations to perform a set of n localized regressions, where n indicates the number of observations. In calibrating a GWR model, the optimal bandwidth, measured as distance or k number of nearest neighbours (KNN), is derived. This KNN is the sample size for each localized regression. MGWR adds to this framework by relaxing the assumption that all processes operate at the same scale, i.e. the bandwidth, or sample size, is allowed to vary for each parameter (Fotheringham et al., 2017). This provides insights into the different process scales that may be present over space. The data set used for the analysis was provided by the Nederlandse Vereniging van Makelaars (NVM), and includes 3632 observations of house transactions in the city of Rotterdam in 2018. The research questions are as follows:

What is the degree of spatial heterogeneity in the Rotterdam housing market? Are there differences with regards to the spatial scale at which processes explaining housing prices operate?

In this study the performance of the MGWR is benchmarked against a hedonic model with locational fixed effects, and a regular GWR model, to explore the potential improvement in accuracy when the spatial scales are correctly specified. The further structure of this paper is as follows. First, in Section 2, the literature with regards to spatial analysis and the econometric issues that may occur in spatial datasets is reviewed, in particular the issues of spatial dependence and heterogeneity. Here we will also discuss previous applications of the (M)GWR model in the literature. Section 3 introduces the data set, which is obtained from the Nederlandse Vereniging van Makelaars en Taxateurs in onroerende goederen (NVM), the collective organisation for real estate agents in the Dutch housing market. We also rationalize the data transformation process that is required for proper calibration of a (M)GWR model. Section 4 introduces the MGWR model, and sheds more light on the bandwidth optimization method. Here we also discuss the kernel function that is used for assigning weights to each observation in the local regressions. Finally, the full results are presented in Section 5, after which a discussion of the limitations and implications occurs in Section 6. One of the major findings of this study is that the variables seem to operate on different spatial scales, i.e. when the number of nearest neighbours included in each local regression is allowed to vary for each parameter, the model fit is highest. This implies that spatial dependence and heterogeneity in data can not simply be captured using locational fixed effects, because fixed effects (and global models in general) assume the same spatial scale for each parameter.

2 Review of Literature

2.1 Spatial Dependence and Spatial Heterogeneity

As mentioned in the introduction, spatial issues are not always correctly addressed in hedonic (real estate) applications. This can be attributed to the presence of spatial dependence and spatial heterogeneity in data, respectively also referred to as second order and first order variation of spatial processes. With spatial dependence, input variables or the respondent variable may be affected by values in a neighboring region. It is likely to occur in housing data sets because nearby properties often show similar structural features and have access to the same amenities, and thus are likely to be spatially dependent (Basu & Thibodeau, 1998; Bourassa et al., 2010). The assumption of independence in the context of OLS is therefore not likely to be valid. Regions may be correlated with neighbors in one of three ways: 1) The value of y in a region might be correlated with the value of y in a neighboring region (lag y); 2) The value of Xs in a region might be correlated with the value of y in a neighboring region (lag X); 3) The residuals ϵ might be related to the residuals in a neighboring region (spatial autocorrelation). The Manski model, also referred to as the general nesting spatial model (Elhorst, 2014), incorporates all three types of correlation in one model:

$$y = \rho W y + \alpha \iota_N + X\beta + WX\theta + \upsilon \tag{1}$$

$$\upsilon = \lambda W \upsilon + \epsilon \tag{2}$$

In this model, ρ is some multiplier, generally <1, Wy are the neighboring values of y, X is y's own vector of independent variables, WX indicate the neighboring values of X multiplied by θ, which is a vector of a lot of different slope parameters. υ is the unexplained variation, which is a function of the neighbor's unexplained variation + ϵ, the error term.

The Manski model in Equation 1 is however not properly equipped to deal with the other concern in spatial models, which is spatial heterogeneity. Although it accounts for potential spatial autocorrelation in data through spatial lags, the model creates parameters that represent an average of the parameter over all locations (this is what we refer to as a stationary coefficient model). Stationary coefficient models should not be applied to analyses of housing prices, because the supply of specific housing characteristics often demonstrates strong spatial patterns in urban areas, and the relationship between explanatory variables and the output variable may therefore differ substantially over regions (Bitter et al., 2007). This can be demonstrated as follows: When household preferences for specific housing characteristics and locational attributes change, this could result in spatial disequilibrium of supply and demand. Greater competition for housing attributes in a specific area should result in higher marginal prices, which in turn may result in spatial heterogeneity. Following this logic, the following hypothesis is derived to explore the extent of spatial heterogeneity Rotterdam:

Hypothesis 1 : *The relationship between the respondent and explanatory variables is not constant over space, i.e. there is spatial heterogeneity in the Rotterdam housing market.*

Spatial heterogeneity can be modelled in both a discrete and a continuous manner. For the discrete approach we implement predefined spatial units in the model, as some sort of location fixed effects. The underlying assumption here is that the spatial range of the unobserved heterogeneity or dependence neatly follows the borders of the spatial unit. In practice, this assumption does not often hold, as omitted neighborhood variables are often of a more complex nature, meaning that the residual spatial autocorrelation can not fully be removed using a fixed-effects approach (Anselin & Lozano-Gracia, 2009). The continuous method does not rely on exogenous assumptions, but rather on a locational attribute, such as a coordinate reference, that is present in each observation. It is therefore better suited to deal with data in which spatial heterogeneity is likely to be present. One method that addresses both spatial dependence and uses a continuous approach to model spatial heterogeneity is the spatial expansion method, which allows parameters to drift based on their location, in an OLS framework (Cassetti, 1972). This is done through interaction of the house characteristics or attributes and the locational variables. However, in a fairly recent study on the Tucson, Arizona housing market, the spatial expansion method is outperformed by another method (Bitter et al., 2007), the geographically weighted regression (GWR), which will be thoroughly discussed in the next section.

2.2 (Multiscale) Geographically Weighted Regression

The GWR allows unique coefficients to be estimated at each location (i.e. a sale point), so it essentially specifies separate regression models for each point of observation. The sample size for the local regressions is referred to as the bandwidth. Each observation within the bandwidth has a weight that is adjusted based on their distance to a specific sale point (i.e. the house under observation), where closer observations are assigned a larger weight. This is further explained in the Methodology section of this paper. Not only does GWR identify spatial heterogeneity in processes and take advantage of spatial dependence in data (Fotheringham et al., 2017), it also does so better than other methods, such as the spatial expansion method mentioned in Section 2.1. Another comparison of GWR was carried out by McCluskey et al. (2013), who compared the prediction accuracy of several mass appraisal methods. Their measures for relative performance of the models are based on the ease with which a model can be explained, the consistency of the model structure, transparency of the output, and the explicitness of the locational element. The study shows that geostatistical localised regression, or GWR as we call it, is superior in terms of reliability, accuracy and explanation, with an artificial neural network (ANN) model following in close pursuit. A study by Farber & Yates (2006) on the Canadian housing market also shows strong support for the superiority of this model, and provides evidence for spatial heterogeneity in several housing attributes. Because of this superiority, the GWR model is used in this research as the standard for estimating the value of a property. In line with Yeh & Hsu (2018), the hedonic pricing model is used as a benchmark against which the performance can be measured.

The single optimal bandwidth that is derived in a GWR can be regarded as a weighted average of the different scales of spatial heterogeneity, where the weighting is assigned based on the explanatory power of each relationship in the local model. This is an apparent shortcoming of GWR, as it can lead to misspecification errors and biased parameter estimates if it is applied to data where multiple spatial scales are at work in the background. This is one of the main advantages of using MGWR instead. MGWR is a type of GWR that does not have the strict assumption that all processes operate at the same scale. In other words, a MGWR model allows the sample size for the localized regressions to differ per variable. It is suggested to be used as the standard model in any analysis where GWR would also be applicable (Oshan et al., 2019). As this method has only recently been developed, it has not yet been applied to the real estate market and could therefore provide some very interesting and novel insights. There is one notable contribution with regards to different spatial scales in the literature, belonging to Helbich et al. (2014). In an attempt to account for the difference between global and local processes, they employ a mixed geographically weighted regression to the Austrian housing market. The model is called mixed in the sense that some of the coefficients are spatially fixed (because of a suspected global process occurring), whereas a number of local processes are allowed to vary spatially, albeit all at the same bandwidth. The mixed GWR provides evidence for significant spatial variation for some of the covariates, whereas policy-based linkages and economic interconnectedness of the housing market are demonstrated in the global effects. The authors of this work rightfully state that there is no empirical consensus on which predictors enter the model globally and which locally. Since no literature is available on the spatial scales of variables, we will use logic to derive the hypotheses.

The following variables are expected to operate on a local scale[1]: a) *MAINTOUT* (the state of maintenance of the house exterior), because it can be argued that the state of surrounding houses can impact the price of a house. Houses in the direct vicinity that are well maintained will probably positively affect the object's sales price, whereas well-maintained houses that are not nearby are not likely to affect the price. b) *GARDEN* (binary variable) is likely to affect the house price in a similar fashion, as the presence of gardens in a neighbourhood is likely to generate positive spill-overs for other houses because of the added space and green in between houses.[2] c) Building period (*BP*) can be correlated to the aesthetics of the house, and therefore might generate positive (or negative) spill-overs in a similar manner to *GARDEN*. In summary, the variables that are expected to be local all have to do with features that can be observed in the direct vicinity of the house, and could therefore generate spill-overs of some sort.

For the following variables we can make the case that they are likely to operate on a global scale: a) *M2* (the lot size in squared meters), as buyers can reasonably compare houses of similar sizes in several neighbourhoods, not necessarily close or adjacent. I.e., it can be reasoned that individuals looking for a house will beforehand have certain expectations or desires with regards to the lot size. b) *CEILING* (the height of the ceiling in meters) follows the same logic, if potential buyers are looking for

[1]For a full overview of the variables and their description see Appendix A1.

[2]Note that this variable refers to private gardens and not public green spaces.

a house with specific dimensions they can reasonably be expected to do so over different areas. c) *ISOL* (the degree of isolation in a house). Although this is probably strongly correlated with the building period of a house, house owners can -depending on their preference or expected utility- go through great lengths to acquire decent or sustainable isolation. Potential buyers can -based on the same preferences- decide to compare houses with similar degrees of isolation over larger distances. Variables *nROOMS*, *CAT*, and *nFLOORS* follow similar strains of reasoning as *M2* and *CEILING* and are therefore expected to be global. The above expectations are expressed in the following hypothesis:

Hypothesis 2 : *The optimal bandwidth (k number of nearest neighbours) is not the same for all parameters, i.e. some processes are local whereas others are global.*

3 Data

3.1 NVM Transaction Data

The data for this research is provided by the Nederlandse Vereniging voor Makelaars (NVM), which is the collective organization for real estate brokers and appraisers in the Netherlands. According to their website, an approximate 75% of Dutch houses is sold by real estate brokers belonging to this organization. Every house transaction that is overseen by a member of this organization is added to the database, along with an assessment of the different characteristics of properties and the transaction prices. Using QGIS software and the CBS Wijk- en Buurtkaart 2018 data we add the neighbourhood that each observation is in, to be able to control for locational fixed effects in the global hedonic model that is used as a benchmark. We can either include this category with around 15 factors or with around 65 factors. In line with Bourassa et al. (2010), who state that increasing the number of submarkets tends to improve the results, the latter option is chosen. Because of the intensive computational requirements of the MGWR model, the data set is reduced to only include houses sold in 2018 in Rotterdam, a major port city in the Randstad area of the Netherlands with a population of 620,000 inhabitants. This results in a total of 3719 observations. For

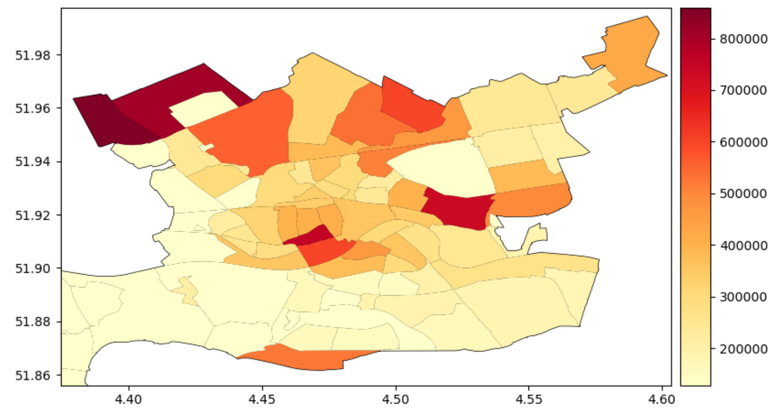


Figure 1: Mean Transaction Price (in €) per Neighbourhood in Rotterdam

the purpose of this exploratory research, i.e. to determine if there are differences in the spatial scales over which valuation processes operate, this sample should be sufficient.

3.2 Variables and the Modeling Environment

The dependent variable is transaction price, which shows a substantial amount of variation across the city (Figure 1). It must be noted that the amount of observations n differs substantially per neighbourhood, and the figure should not be interpreted as an accurate representation of the average value of houses. Its sole purpose is to illustrate the spatial variation. An overview of the included variables can be found in Table A1. To correct for potential skewing, the dependent variable underwent a logarithmic transformation. Variables to do with a house's structure are the lot size (*M2*), the height of the ceiling (*Ceiling*), state of interior and exterior maintenance (*MaintIn and MaintOut*, the degree of isolation (*Isol*), the number of floors (*nFloors*), number of rooms(*nRooms*), and presence of a garden (*Garden*). Also included are variables indicating whether the unit under observation is a house or apartment (*CAT*), the building period (*BP*), locational aspects (*LOC*), and the neighbourhood each observation is in as a fixed effect (*NBH*).

As the NVM dataset only includes the Postal Code and Address as locational attributes, the data had to be further prepared for use in a MGWR model. More specifically, using the abovementioned locational attributes all observations were geocoded

Table 1: Descriptive Statistics

	Mean	Std. Dev.	Min.	Max.
Transaction price	314,595	213,093	51,000	2,900,000
M2	106.79	46.18	27.0	517.0
CEILING	3.01	0.34	2.10	5.86
MAINTIN	6.77	1.18	1	9
MAINTOUT	6.89	0.90	1	9
ISOL	1.91	1.73	0	5
nFLOORS	1.87	0.95	1	6
nROOMS	3.98	1.52	1	16
GARDEN	0.34	0.47	0	1
CAT	0.68	0.47	0	1
BP_1	0.39	0.49	0	1
BP_2	0.27	0.44	0	1
LOC_2	0.15	0.35	0	1
LOC_4	0.17	0.38	0	1

to longitude / latitude coordinates using an external API. Furthermore, a number of observations that indicated a lot size of 0 squared meters were dropped, as they are likely to be errors and will obscure the results. This left a total number of observations of 3632. Also, because MGWR does not allow for inclusion or interpretation of nominal categorical variables, *BP*, *LOC* and *NBH* were one hot encoded to perform binarization of these categories. The categorical variables with a natural ordinal relationship (e.g. *MaintIn* and *Isol*) are treated as continuous, and therefore one hot encoding is not required. Finally, for the MGWR model, both the continuous X variables and the Y variable are standardized so they are centered at zero, and are based on the same range of variation. This allows for comparison of each of the bandwidths that is obtained (Oshan et al., 2019).

4 Methodology

The MGWR module provides parameter estimates for the global (OLS) and local (GWR / MGWR) hedonic house price models. Both will be briefly introduced in this section.

4.1 Global Hedonic Model

In a hedonic price function, housing price is expressed as a function of a number of housing characteristics. In the results section of this paper the Ordinary Least Squares (OLS) model is referred to as the global hedonic model (GHM), because each coefficient is assumed to operate at the scale of the entire data set. Although previous works have already demonstrated the superiority of GWR as compared to global models (Farber & Yates, 2006; Bitter et al., 2007; McCluskey et al., 2013), the GHM is included here so we can interpret the results of the local models relative to the global model. See below the GHM in formula,

$$y_i = \beta_0 + \sum_{k=0}^{m} \beta_k x_{ik} + \epsilon_i \tag{3}$$

where x_{ik} can be read as the ith observation of the kth independent variable. Note that a locational element is not yet explicitly present in the observations, location is only incorporated in the global hedonic model in the form of a fixed-effects dummy of the neighbourhood, and is thus part of the vector x_{ik}. The GWR model itself is rather similar to the above model, except for the fact that it performs parameter estimation at each observation point, as is demonstrated in the next section.

4.2 (M)GWR

4.2.1 GWR & Spatial Weights Matrix

Whereas the GHM offers the same set of coefficients for all observations over a geographical area, the Geographically Weighted Regression (GWR) produces a separate regression, and therefore a separate set of coefficients, for each of the observations. Each of these local regressions only incorporates those neighbours that are within the obtained bandwidth. The optimal bandwidth is selected by optimizing a model fit criterion, such as the R^2 or the corrected Akaike Information Criterion (AICc). The latter one is suggested by Oshan et al. (2019) (and also applied in this work) because it penalizes smaller bandwidths, which result in models that are more complex and consume more degrees of freedom.

The obtained bandwidth can either be expressed in Euclidean distance or as in the current work, the k number of nearest neighbours, indicating the range of the local data-borrowing scheme. This bandwidth is important, because it indicates the scale at which a variable operates. If the optimal bandwidth approaches n, we can say that the variable operates on a global scale, i.e. almost all other observations are considered in the local regression, albeit with different weights through a distance-decay weighting function. If the optimal bandwidth is relatively low, i.e. just a small amount of neighbours have non-zero values in the spatial weights matrix, we can say that the variable operates on a relatively local scale.

GWR is equipped to take into account both the spatial dependence and non-stationarity that are discussed as issues in the literature section. In formula, the GWR model looks as follows:

$$y_i = \beta_{i0}(u_i, v_i) + \sum_{k=0}^{m} \beta_k(u_i, v_i)x_{ik} + \epsilon_i \tag{4}$$

The term $\beta_k(u_i, v_i)$ represents the coefficient of variable k at point i. The set of longitude / langitude coordinates corresponding to each observation is expressed in (u_i, v_i) The parameters of a GWR model can be estimated using the following function, where X represents the design matrix and $W(u_i, v_i)$ signifies the $n * n$ spatial weights matrix:

$$\beta(u_i, v_i) = (X^T W(u_i, v_i)X)^{-1} X^T W(u_i, v_i)y \tag{5}$$

In the model, a spatial weight matrix is obtained through use of a kernel function that emphasises those observations within the bandwidth range that are closer to the calibration point, and assigns less weight to those further away. Three kernel functions that can be used for this are the Gaussian, exponential, and bi-square functions, each of which can have a fixed or an adaptive weighting process. Fixed kernels use a constant distance, meaning that the number of nearest neighbours varies if the data is not uniformly distributed. This also implies there can be calibration issues in sparsely populated regions. The adaptive kernel allows the distance to vary while ensuring that each local regression uses the same number of nearest neighbours. In this study, an adaptive bi-square kernel is applied to derive the spatial weights matrix:

$$W_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}}{h_i}\right)^2\right]^2, & \text{if } d_{ij} < h_i \\ 0, & \text{if } d_{ij} > h_i \end{cases} \tag{6}$$

The main reason for this is that for both the Gaussian and exponential kernels all observations retain non-zero weight, so that even observations that are further away remain of influence (Oshan et al., 2019). For the bi-square kernel, each observation beyond the distance or bandwidth threshold has weight 0. In effect, for a bi-square kernel function the derived bandwidth can be interpreted as the distance or number of k nearest neighbours beyond which the remaining observations have no influence. In equation 6 the term in between the square brackets shows the distance decay function. Notice that the weight of the observation is reduced as the distance between location i and j (d_{ij}) increases. If the distance d_{ij} is larger than the bandwidth at location i (h_i), the assigned weighting for that observation is equal to 0, indicating that the observation is not included in the local regression. Once the kernel function and bandwidth selection method have been decided, a GWR model can be calibrated.

4.2.2 MGWR and Spatial Scale Variation

Even though GWR appears to be superior to other available methods for real estate appraisal, it still operates under the assumption that the bandwidth is the same for all of the parameters. In instances where there are multiple spatial scales present, some of those scales are bound to be misspecified and may result in biased estimates. MGWR provides a solution for this, by allowing each variable to have a distinct scale. This is expressed in the following function:

$$y_i = \beta_{0i}(u_i, v_i) + \sum_{k=0}^{m} \beta_{bwk}(u_i, v_i)x_{ik} + \epsilon_i \tag{7}$$

Equation 7 looks similar to the GWR model in Equation 4, but includes the term β_{bwk}, indicating the bandwidth for calibration of the k^{th} relationship. Different bandwidths would imply that the relationships between the explanatory variables and the log of price on any given location have different spatial weighting matrices. For ease of reference, we will refer to processes as local if the optimal derived bandwidth is within the 5th percentile of the total sample size, i.e. smaller than 180. We distinguish the processes as global if they appear to operate on the scale of the total sample size, i.e. close to 3632. Standardization of the variables (a Z-transformation so each variable has $mean = 0$ and $standard\ deviation = 1$) allows the bandwidths to be interpreted as direct indicators of the spatial scale of the relationship between each explanatory variable and the log of price. If standardization is not applied the bandwidths also reflect the scale and variation in each variable, but it does make the results easier to interpret (Oshan et al., 2019).

All three models described in this chapter are applied to the data, in order to compare the outcomes and model fits. The global hedonic model is the only model that includes locational fixed effects (variable *NBH*), so we can observe differences between the discreet (GHM) and continuous (GWR and MGWR) approach to modelling spatial heterogeneity. It can be argued that the neighbourhood a house belongs to affects the housing price for reasons other than merely location, as the reputation of neighbourhoods (even those that share borders) may be very different. Therefore, even though these fixed effects may not be sufficient to fully capture spatial heterogeneity in the data, they may capture other unobserved effects. It would therefore make sense to include them in the GWR and MGWR models as well. However, this is not done for several reasons. First of all, given the fact that each observation can only belong to one (of approximately 65) neighbourhoods, the localized models are very prone to singular matrices (or multi-collinearity) as there may be no variation for local subsets with a lower bandwidth. Second of all, the backfitting algorithm that retrieves the optimal bandwidth per variable would likely be running for several days to run a single model. Although this is not a problem in itself, the principal objective of this research is to demonstrate the spatial scales of processes, and not necessarily to achieve a model with the highest R^2. Although that can quite possibly also be achieved using MGWR, the author leaves that for future research.

With regard to GWR and MGWR, several models are calibrated, using a different number of variables each to explore the robustness of the bandwidth selection procedure. Unfortunately, due to technical constraints, two of the variables (CAT and BP_2) can only be included in the MGWR and not in the GWR. This is because the iteration process that searches for the optimal bandwidth(s) has a starting value that can not be altered for the GWR. For some of the binary variables, especially those with little variability, this may lead to multicollinearity issues in lower bandwidths. MGWR has a way to circumvent this problem, by allowing the user to specify an initial search value.

5 Results

To study the first hypothesis regarding spatial heterogeneity, we discuss the spatial patterns that are captured in the maps in Figure 2 and Figure 3. Furthermore we compare the model fits as shown in the bottom of Table 2 as well as the mean, minimum and maximum coefficient values per variable as displayed in Table 3. To provide evidence for the second hypothesis on spatial scales we discuss the bandwidths that are displayed in Table 2 as well as the colour maps in Figure 3.

5.1 Interpretation of MGWR and Spatial Heterogeneity

Although it was almost certain that we would find spatial heterogeneity based on the nature of real estate markets and the uniqueness of properties under observation, it is interesting to see the patterns that emerge. First of all, if we compare the GWR-

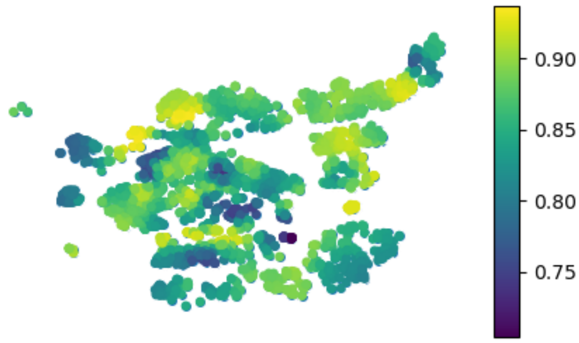


Figure 2: Local R2 statistics for GWR-11s Model (n = 3632)

11s model (the number indicates the number of variables included, with *s* signifying that the variables are standardized) with the GHMs model we see that the model fit improves if the relationships are allowed to vary over space (Table 2), as opposed to the use of only locational fixed effects. For the GWR-11s model in specific, we can see in Figure 2 that the explanatory power of the model differs over space, with some areas reaching an R^2 statistic of over 0.90 where for areas with the lowest fit this value reaches no more than 0.75. Within each variable there is also a fairly large dispersion between the size of the coefficients, as is observed in the minimum and maximum values of the coefficients in Table 3. Per illustration, Figure 3 shows the distribution of coefficients for each local regression over space.

We will briefly discuss the sign and magnitudes of the coefficients.[3] *M2* is positive and mainly significant throughout the city, although the magnitude of the coefficients takes on values between 0.13 and 0.96 on a 1-point scale, signalling a large degree of heterogeneity. There are no clear patterns or clusters that form. The spread of coefficients for *CEILING* is more modest, with estimates between approximately 0.05 and 0.09. A clear pattern emerges, with the effect being smaller in the northeast than in the rest of the city. The presence of this cluster is an indicator of spatial dependence. *MAINTIN* fluctuates slightly, with negative outliers in the centre of the city, amongst a cluster of non-significant (grey) coefficient estimates. *MAINTOUT* has a higher

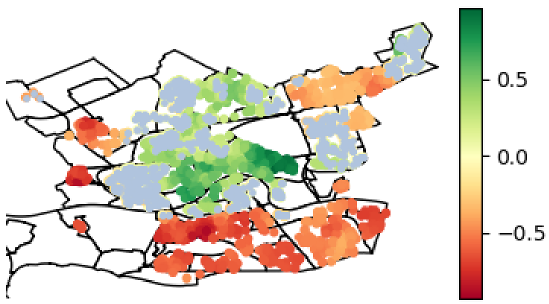
[3]Unfortunately, we have no data on demographics per neighbourhood. More research will have to point out how the effects can be related to demographic information such as mean income or average household size.

dispersion but the effect is not statistically significant in nearly all local regressions. *ISOL* has a slight variation in coefficient estimates, with higher values in the west and center of the city than in the east. The number of floors ($nFLOORS$) is negatively related to housing price in all locations, although it appears that about half of the observations are not significant at the .05 level. This relationship is relatively stable over space. The number of rooms shows spatial clustering with the positive relationship gradually becoming larger as we move from east to west. The effect of having a garden is positive, as could be expected, with the effect being fairly uniform over space. The parameter estimates in the east are not statistically significant. Throughout the city, having a house as opposed to having an apartment yields a higher selling price, ceteris paribus, although the effect is not significant in the center. This might be because the bandwidth is relatively small (342) and there are not many houses in the centre of Rotterdam. Both dummy variables for the building period (BP_1 and BP_2) show mainly non-significant parameter estimates, although buildings built after 1990 (BP_2) appear to yield a higher sales price, ceteris paribus, compared to the reference category, which represents houses built before 1945. This effect is predominant in the southeast of the city. Adjacency to water (LOC_1) is positively correlated to transaction price across the city, with higher coefficient values in the neighbourhood of Hillegersberg in the centre north. The effect of having a free view LOC_2 is positive and stationary over space, with non-significant values in the east.

Due to the apparent nonstationarity of many parameters as described above, Hypothesis 1 is not rejected, as there is clear evidence of spatial heterogeneity in the data.

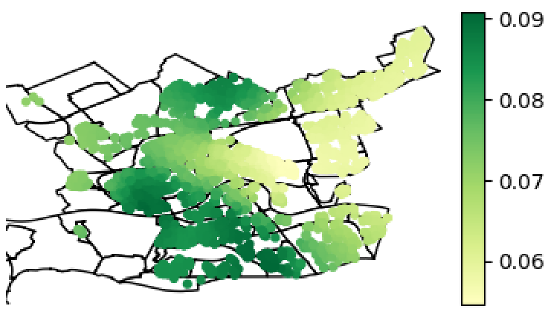
5.2 Spatial Scale of Parameters

The main objective of this paper was to explore the spatial scale of processes. The second hypothesis states that the optimal bandwidth, measured as the k number of nearest neighbours, is not the same for all parameters. Support for this hypothesis can be found in Table 2, where the GWR with fixed bandwidths is compared to the MGWR model that allows spatial variation between variables. We observe that in the MGWR model the AICc is lower and the adjusted R^2 is higher relative to the GWR



(a) Intercept (BW = 43)

(b) M2 (BW = 46)

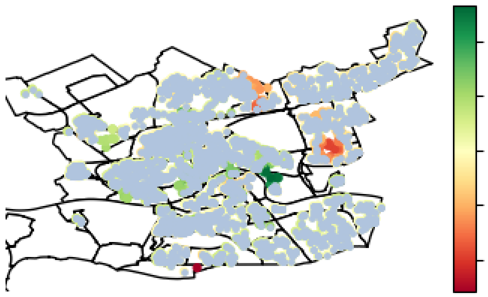


(c) Ceiling (BW = 1291)

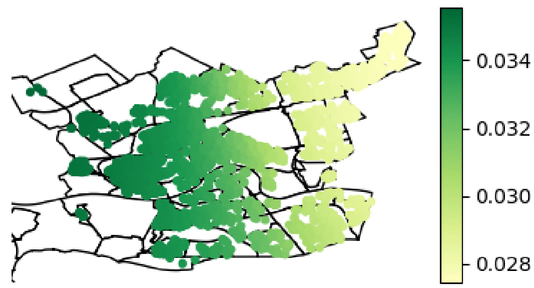
(d) MaintIn (BW = 642)

model. This suggests that the model with spatial variation over parameters (MGWR) best represents the data. Applying the same bandwidth to all variables therefore likely introduces bias to the model, as can be observed through the difference in coefficients between both models.

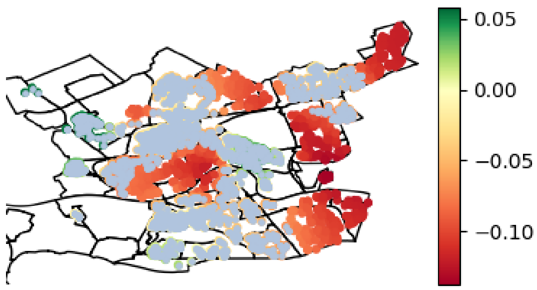
The results provide us with unique insights into the locality or globality of processes. At the local level (meaning only the closest observations are included in the distance-decay weighting scheme) we encounter the variables lot size (in M2) with a bandwidth of 46, and the state of maintenance of the house exterior ($MAINTOUT$) with a bandwidth of 57. For the latter, we hypothesized that the effect would indeed be local because of local spillovers due to aesthetics. However, the predominant grey color shows that in almost all of the cases the local coefficient estimates are not statistically significant for this variable. The outcome for lot size ($M2$) is more surprising, as we expected this



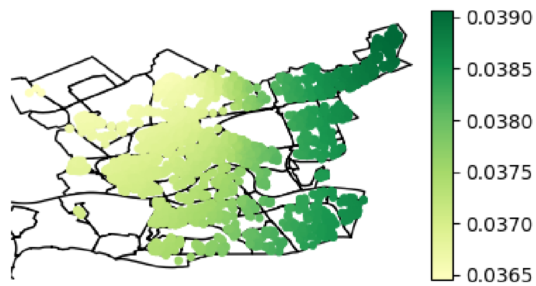
(e) MaintOut (BW = 57)



(f) Isol (BW = 3626)



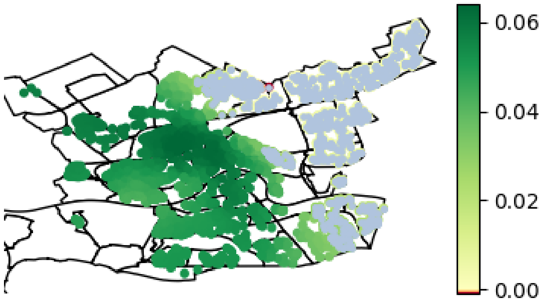
(g) nFloors (BW = 383)



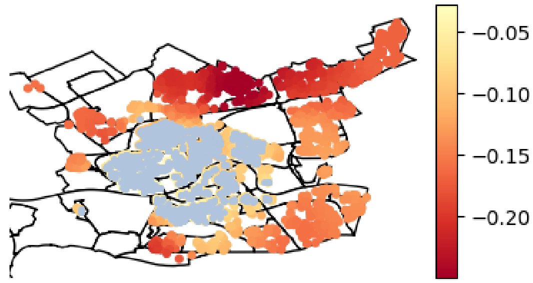
(h) nRooms (BW = 3632)

to operate on a global scale. One possible explanation for the locality of *M2* is that an individual's utility could be based not on absolute, but on relative comparisons. In this instance, it could mean that people care more about lot size relative to the lot size of the closest neighbouring houses. This notion is supported by earlier works such as a thought experiment by Frank (2005) where individuals favour a situation with a smaller absolute but a larger relative house size. A recent study by Bellet (2019) provides further empirical evidence on the existence of positional externalities in housing size.

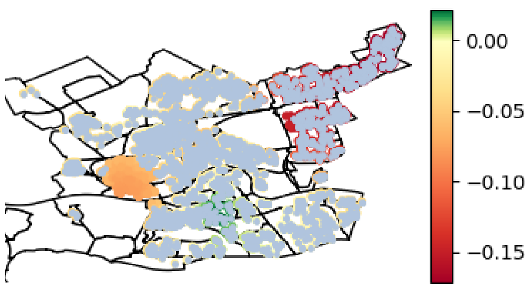
Variables that appear to have a global effect are the degree of isolation, the number of rooms, and whether the house has a free view (LOC_2). For these variables a GHM is therefore likely to perform similarly to a (M)GWR. It makes sense for the first two variables to be global, because it is unlikely that the degree of isolation or number of rooms of only nearby houses affect the price. Rather, it is logical that house owners or



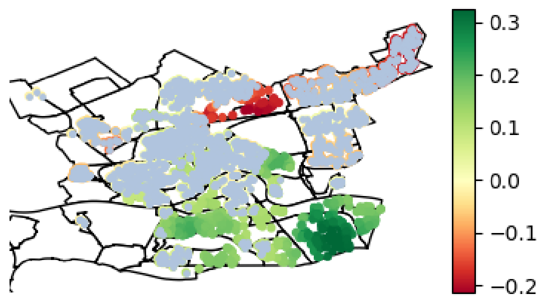
(i) Garden (BW = 1249)



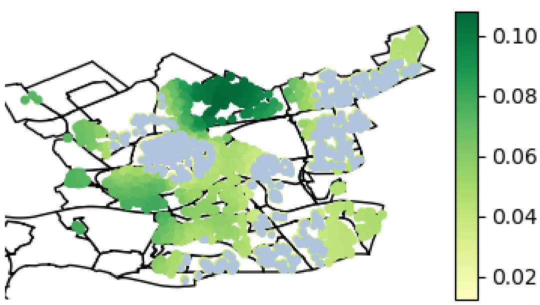
(j) CAT (BW = 342)



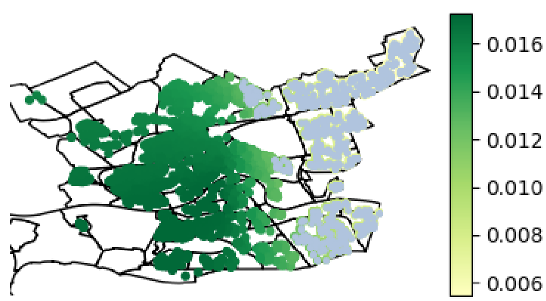
(k) BP_1 (BW = 517)



(l) BP_2 (BW = 156)



(m) Loc_1 (BW = 644)



(n) Loc_2 (BW = 3626)

Figure 3: MGWR Parameter Estimates and Bandwidth

buyers consider houses with similar features over a wider area of space, as the degree of isolation and number of rooms can not be observed from outside and are therefore

Table 2: Comparison GWR-11s & MGWR-13s

	GHMs	GWR-11s		MGWR-13s	
Variable	Coefficient	Coefficient	Bandwidth	Coefficient	Bandwidth
Intercept		−0.005	176	−0.022	43
M2	0.541***	0.560	176	0.508	46
CEILING	0.088***	0.092	176	0.074	1291
MAINTIN	0.085***	0.069	176	0.092	642
MAINTOUT	0.034***	0.037	176	0.024	57
ISOL	0.024***	0.049	176	0.032	3626
nFLOORS	−0.090***	−0.034	176	−0.052	383
nROOMS	0.040***	0.042	176	0.038	3632
GARDEN	0.035***	0.059	176	0.040	1249
CAT	−0.151***			−0.121	342
BP_1	−0.042***	−0.061	176	−0.051	517
BP_2	0.058***			0.033	156
LOC_1	0.072***	0.067	176	0.050	644
LOC_2	0.016***	0.013	176	0.014	3364
NBH	β_{14-77}				
ENP		575		613	
AICc		1982		1297	
Adj. R^2	0.891	0.913		0.933	

*The GHM model is standardized. *** indicates that the coefficient is significant at the 0.01 level.*

unlikely to generate local spillovers. With regards to houses having a free view,

Future research will have to point out if these scales and relationships hold over a larger area, and if so what might be the underlying cause(s).

Interpretation of the intercept doesn't make sense, as the variables are all standardized. The standardization also makes it harder to interpret the exact effect of each variable on the transaction price. However, because all variables now operate on

Table 3: MGWR-13s Coefficient Estimates

	Coefficient				
Variable	Mean	Std. Dev.	Min.	Max.	Bandwidth
Intercept	−0.022	0.451	−0.924	0.964	43
M2	0.508	0.120	0.135	0.954	46
CEILING	0.074	0.120	0.055	0.091	1291
MAINTIN	0.092	0.039	−0.060	0.139	642
MAINTOUT	0.024	0.093	−0.516	0.532	57
ISOL	0.032	0.002	0.027	0.036	3626
nFLOORS	−0.052	0.047	−0.137	0.058	383
nROOMS	0.038	0.001	0.036	0.039	3632
GARDEN	0.040	0.021	−0.001	0.064	1249
CAT	−0.121	0.055	−0.249	−0.027	342
BP_1	−0.051	0.052	−0.171	0.022	517
BP_2	0.033	0.112	−0.213	0.326	156
LOC_1	0.050	0.021	0.013	0.108	644
LOC_2	0.014	0.004	0.005	0.017	3364

the same scale we can deduce the relative size of each of their effects, where values closer to 1 indicate a larger effect. Not surprisingly, the size of the residence (*M2*) has -on average- the largest role in explaining housing prices, even though this differs substantially over space. Table 3 shows the distribution of each effect over all 3633 local regressions, and is another stellar demonstration of spatial heterogeneity: We observe relatively large differences between the minimum and maximum regression coefficients for most variables, as is also reflected in the relatively high standard deviations.

5.3 Robustness Checks

To check how robust the results regarding the spatial scales are, the bandwidth selection procedure is run for several models containing different numbers of variables (Table 4).

Table 4: Bandwidths and Model Statistics for Different (M)GWR Calibrations

Model	MGWR-8s	MGWR-8u	GWR-11s	MGWR-11s	MGWR-13s
Intercept	43	129	176	43	43
M2	43	43	176	45	46
CEILING	3608	508	176	1212	1291
MAINTIN	642	1243	176	642	642
MAINTOUT	48	307	176	47	57
ISOL	114	458	176	387	3626
nFLOORS	381	420	176	343	383
nROOMS	3623	3632	176	3632	3632
GARDEN	343	263	176	367	1249
CAT					342
BP_1			176	1005	517
BP_2					156
LOC_1			176	502	642
LOC_2			176	2306	3364
ENP	649	291	575	629	613
AICc	1665	-2177	1982	1557	1297
Adj. R^2	0.927	0.9	0.913	0.929	0.933

The number behind the model represents the number of variables used, where the last letter indicates a model with standardized (s) or unstandardized (u) variables. ENP indicates the effective number of parameters, and AICc is the corrected Akaike Information Criterion.

Overall, the optimal bandwidths for most variables vary only marginally. However, inclusion of the category (CAT) and BP_2 variable yields a substantial increase in the optimal bandwidth for ISOL, GARDEN and LOC_2, and it results in a decrease for BP_1. A possible and logical explanation for this is that the models that do not include these variables suffer a form of omitted variable bias. The case can be made that houses are more likely to have a garden than apartments. Furthermore, changing policies over the years may result in more houses being constructed in one period, and more apartments in another. The sharp increase for isolation is odd, but can be tied to the better specification of building periods, as more recently constructed residences tend to have different (often better) isolation than older buildings.

6 Conclusion

6.1 Discussion and Implications

This study set out to investigate the spatial scales and the degree of spatial heterogeneity present in the Rotterdam real estate market. The problem we intended to tackle is formulated by Helbich et al. (2014), who state that there is an absence of empirical consensus on the locality or globality of processes in hedonic house price models. I demonstrated the different spatial scales at which processes determining housing prices operate with help of a MGWR model. Through use of AICc optimization, different bandwidths were found to be optimal for different variables. This suggests that spatial heterogeneity in housing price markets can not simply be captured using locational dummy variables, and that models determining housing price will have to be adjusted accordingly. The MGWR model developed by Fotheringham et al. (2017) proves to be an interesting and effective approach in tackling the problems of spatial dependence and heterogeneity.

The significance and relevance of the results is two-fold. Firstly, the MGWR model in a hedonic house price context appears to be a better specified, and more accurate model than the GWR model, which on repeated occasions proved to be superior to several other methods. The current work -to the best of my knowledge- is the first hedonic house price application of a MGWR. This apparent superiority indicates that practitioners of real estate appraisal, both in the academic world as in the field, stand to benefit from use of this model. The second implication of this research relates to policy. Through estimation of the bandwidths for each variable, we gain insights into the scale at which processes operate. This means that parties with a vested interest in elevating the value of properties in a certain neighbourhood (for instance a local government in an attempt to increase welfare locally) can more efficiently specify investment targets. In particular, using knowledge on the scale of processes, whether they be local or global in nature, the extent of spatial spillovers from investment can be effectively evaluated.

6.2 Limitations and Future Research

It is the author's belief that MGWR has a lot of potential for applications within the field of real estate appraisal, and the current work has only scratched the surface. There are some limitations and issues that are not touched upon in this work, while they merit future research. First of all, MGWR does not allow the inclusion of categorical variables, so each category would have to be added as a separate binary variable. For some of the variables this was done (LOC and BP to be precise), but other variables, for instance *NBH*, representing one of 65 different neighbourhoods, were deliberately left out to maintain a clear interpretability and prevent multicollinearity issues for parameters with lower bandwidths. Furthermore, not all variables that you may want to see in a hedonic housing price application were present in the data set. Such variables include neighbourhood effects, e.g. the rate of unemployment or the proportion of academics in an area. A study by Kestens et al. (2006) suggests that detailed household-profile data helps explain spatial heterogeneity while reducing spatial dependence. The results of their study show that a household's education level plays a significant role, as higher-educated households tend to pay a premium to fulfill social homogeneity. It is possible that the vector of *NBH* variables in the GHM partially captured these effects (therefore potentially rasing the explanatory power of the model) while it was not used in the localised models. Another limitation of the current work is that there is no further distinction with regards to the time of the transaction, other than the year in which it took place (2018). Most likely a model that does include this will explain more of the variation, as heated housing markets may often differ several percentage points between the beginning and end of a year. This was certainly the case for Rotterdam in 2018.

This study did not address or try to explore the out-of-sample prediction accuracy of MGWR in comparison with other methods. Future research will have to demonstrate whether the MGWR model is not only effective for modeling spatial heterogeneity, but if its potential to perform accurate real estate appraisal is superior to that of other advanced statistical techniques. It should also be noted that the MGWR was applied in an urban context only, due to the larger availability of data on similar properties in the city. In theory, MGWR could be applied to more rural areas as well, although the standard errors will likely be higher, with less precise estimates as a result. However, it can be argued that the spatial scale of certain processes may differ between urban and rural areas. In particular, the lack of landmarks or a business centre in rural centres could induce higher bandwidths. Finally, future research could incorporate a larger study sample that includes several cities. This way, the economic interconnectedness of housing markets can be explored for each variable.

References

[1] Anselin, L., Lozano-Gracia, N. (2009). Spatial hedonic models. In: Mills, T.C., Patterson, K. (eds) Palgrave handbook of econometrics. Palgrave Macmillan, London.

[2] Barańska, A. (2013). Real estate mass appraisal in selected countries – Functional systems and proposed solutions. *Real Estate Management and Valuation, 21*(3), 35-42.

[3] Basu, A., & Thibodeau, T.G. (1998). Analysis of spatial autocorrelation in house prices. *Journal of Real Estate Finance and Economics, 17*(1), 61-85.

[4] Bellet, C. (2019). The McMansion effect: Top housing size and positional externalities in U.S. suburbs. *SSRN Electronic Journal.* Available at http://dx.doi.org/10.2139/ssrn.3378131

[5] Bitter, C., Mulligan, G.F., & Dall'erba, S. (2007). Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems, 9*(1), 7-27.

[6] Bourassa, S.C., Cantoni, E., & Hoesli, M. (2010). Predicting house prices with spatial dependence: A comparison of alternative methods. *Journal of Real Estate Research, 32*(2), 139-159.

[7] Brunauer, W.A., Lang, S., Wechselberger, P., & Bienert, S. (2010). Additive hedonic regression models with spatial scaling factors: An application for rents in Vienna. *The Journal of Real Estate Finance and Economics, 41*(4), 390-411.

[8] Burkey, M.L. (2018). Spatial econometrics and GIS YouTube playlist. *Region, 5*(3), R13-R18.

[9] Casetti, E. (1972). Generating models by the expansion method: Applications to geographical research. *Geographic Analyst, 4*(1), 81-91.

[10] Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *The Economic Record, 88*(1), 2-9.

[11] Elhorst, J.P. (2014). *Spatial Econometrics.* Springer, Berlin, Heidelberg.

[12] Farber, S., & Yates, M. (2006). A comparison of localized regression models in a hedonic house price context. *Canadian Journal of Regional Science, 29*(3), 405-420.

[13] Frank, R.H. (2005). Positional externalities cause large and preventable welfare losses. *American Economic Review, 95*(2), 137-141.

[14] Fotheringham, A.S., Yang, W., & Kang, W. (2017). Multiscale Geographically Weighted Regression (MGWR). *Annals of the American Association of Geographers, 107*(6), 1247-1265.

[15] Fik, T.J., Ling, D.C., & Mulligan, G. (2003). Modeling spatial variation in housing prices: A variable interaction approach. *Real Estate Economics, 31*(4), 623-646.

[16] Helbich, M., Brunauer, W., Vaz, E., & Nijkamp, P. (2014). Spatial heterogeneity in hedonic house price models: The case of Austria. *Urban Studies, 51*(2), 390-411.

[17] Hodges, M.B. (1993). Three approaches? *The Appraisal Journal, 61*(4), 553-564.

[18] IAAO International Association of Assessing Officers (2012). *Standard on Mass Appraisal of Real Property*, Kansas City, Missouri, USA.

[19] Kestens, I., Thériault, M., & Des Rosiers, F. (2006). Heterogeneity in hedonic modelling of house prices: Looking at buyers' household profiles. *Journal of Geographic Systems, 8*(1), 61-96.

[20] Kettani, O., & Oral, M. (2015). Designing and implementing a real estate appraisal system: The case of Québec Province, Canada. *Socio-Economic Planning Sciences, 49*(issue), 1-9.

[21] Levitt, S.D., & Syverson, C. (2008). Market distortions when agents are better informed: The value of information in real estate transactions. *The Review of Economics and Statistics, 90*(4), 599-611.

[22] Li, Z., Oshan, T., Fotheringham, S., Kang, W., Wolf, L., Yu, H., & Luo, W. (2019). *MGWR 1.0 User Manual.* Tempe, Arizona: Spatial Analysis Research Center (SPARC) Arizona State University.

[23] Mark, J., & Goldberg, M.A. (1988). Multiple regression analysis and mass assessment: A review of the issues. *The Appraisal Journal, 56*(1), 89-109.

[24] Mason, J.J. (1993). Under the microscope: The cost approach. *The Appraisal Journal, 61*(1), 116-128.

[25] Matthews, N. (1921). The valuation of property in the Roman law. *Harvard Law Review, 34*(3), 229-259.

[26] McLaren, N., & Shanbhogue, R. (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin*, No. 2011 Q2.

[27] McMillen, D.P., & Redfearn, C.L. (2010). Estimation, interpretation, and hypothesis testing for nonparametric hedonic house price functions. *Journal of Regional Science, 50*(3), 712-733.

[28] Nevett, L. (2000). A real estate 'market' in Classical Greece? The example of town housing. *Annual of the British School of Athens, 95*, 329-344.

[29] Oshan, T.M., Li, Z., Kang, W., Wolf, L.J., & Fotheringham, A.S. (2019). MGWR: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *International Journal of Geo-Information, 8*(6), 269-308.

[30] Pagourtzi, E., Assimakopoulos, V. Hatzichristos, T., & French, N. (2003). Real estate appraisal: A review of valuation methods. *Journal of Property Investment & Finance, 21*(4), 383-401.

[31] Peterson, S., & Flanagan, A.B. (2009). Neural network hedonic pricing models in mass real estate appraisal. *Journal of Real Estate Research, 31*(2), 147-164.

[32] Ramsland, M.O., & Markham, D.E. (1998). Market-supported adjustments using multiple regression analysis. *The Appraisal Journal, 66*(2), 181-191.

[33] Van Veldhuizen, S., Vogt, B., & Voogt., B. (2016). Internet searches and transactions on the Dutch housing market. *CPB Discussion Paper 325.*

[34] Varian, H.R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives, 28*(2), 3-28.

[35] Wei, C.H., & Qi, F. (2012). On the estimation and testing of mixed geographically weighted regression models. *Economic Modelling, 29*, 2615-2620.

[36] Wyatt, P.J. (1997). The development of a GIS-based property information system for real estate valuation. *International Journal of Geographical Information Science, 11*(5), 435-450.

[37] Xu, T. (2008). Heterogeneity in housing attribute prices: A study of the interaction behaviour between property specifics, location coordinates and buyers' characteristics. *International Journal of Housing Markets and Analaysis, 1*(2), 166-181.

[38] Yu, H. Fotheringham, A.S., Li, Z., Oshan, T., Kang, W., & Wolf, L.J. (2019). Inference in multiscale geographically weighted regression. *Geographical Analysis, 0*, 1-20.

[39] Yeh, I.C., & Hsu, T.K. (2018). Building real estate valuation models with comparative approach through case-based reasoning. *Applied Soft Computing, 65*, 260-271.

7 Appendices

Table A1: Description of Variables

Variable	Description
Transaction price	Selling price in euros (logarithm of)
M2	Lot size in squared meters
CEILING	Height of the ceiling in metres
MAINTIN	State of interior maintenance
MAINTOUT	State of exterior maintenance
ISOL	Degree of insulation
nFLOORS	Number of floors
nROOMS	Number of rooms
GARDEN	Presence of a (private) garden (1)
CAT	House (0) or apartment (1)
BP	Matrix of binaries indicating building period:
	$<$1945 (BP_1+ $BP_2 = 0$);
	1945 $-$ 1990 ($BP_1 = 1$);
	$>$ 1990 ($BP_2 = 1$)
LOC	Matrix of binaries indicating location:
	None ($LOC_1 + LOC_2 = 0$);
	Next to water ($LOC_1 = 1$
	Free view ($LOC_2 = 1$)
NBH	Matrix of binaries indicating neighbourhood, to be used in the global hedonic model only
LONG / LAT	Longitude and latitude coordinates that are tied to each observation

Due to the recent development of the MGWR model, and the fact that the researchers and developers are still adding features to the model, it can only be used in its original Python environment where it can be imported as a module, or in the Graphic User Interface (GUI) software that was developed by the researchers. I chose the former as it provides more options (e.g. manually setting the bandwidth for a MGWR model), but the software should suffise for general purposes and can be downloaded at no expense. For a better overview of how to implement the code one can contact the author or refer to the MGWR Github page or the accompanying implementation paper by Oshan, Li, Kang, Wolf, and Fotheringham (2019).