Combining Property Price Predictions from Repeat Sales and Spatially Enhanced Hedonic Regressions

Are Oust, Simen N. Hansen, Tobias. R. Pettrém

Abstract

Hedonic regression and repeat sales are commonly used methods in real estate analysis. While the merits of combining these models when constructing house price indices are well documented, no research on the utility of adopting the same approach for residential property valuation has been conducted to date. Specifically, house value estimates were obtained by combining predictions from repeat sales and various hedonic regression specifications, which were enhanced to account for spatial effects. Three of these enhancements—*regression kriging*, *mixed regressive − spatial autoregressive model* and *geographically weighted regression*—are widely utilized spatial econometric models. However, a fourth augmentation, which addresses systematic residual patterns in regressions with district indicator variables and the presence of outliers in housing data, was also proposed. The resulting models were applied to a dataset containing 16,417 real estate transactions in Oslo, Norway, revealing that, when the repeat sales approach is included, it reduces the median absolute percentage error of solely hedonic models by 6.8−9.5%, where greater improvements are associated with less accurate spatial enhancements. These improvements can be attributed to the inclusion of both spatial and non-spatial information inherent in previous sales prices. While the former has limited utility for well-specified spatial models, the non-spatial information implicit in previous sales prices likely captures otherwise difficult to observe phenomena, potentially making its contribution highly valuable in automated valuation models.

Keywords: House prices, Automatic valuation models, Spatial models, Urban economics, Repeated sales, Hedonic regressions, Big data

1. Introduction

Accurate property valuation is essential for reducing the inherent uncertainty in housing transactions. As home purchase represents the largest investment most individuals will ever make in their lifetime, uncertainty tolerance is low. This psychological component of what is essentially a financial transaction arguably underpins the real estate agent industry and its business of human appraisal of property market value (Levin, 2001). Accurate property valuation is also critical for housing research. New, well-specified big datasets combined with more computer power have made it cost effective to construct more accurate valuation models. Corcoran and Liu, 2014 points out that the growing demand for automatically generated housing value estimates, as an efficient and cost-effective alternative, may potentially contribute to a more transparent housing market.

In this article, benefits derived by combining property price predictions yielded by two well-known valuation methods—repeat sales and hedonic regression—were investigated. The developed models were tested by applying them to 16,417 residential property transactions in Oslo, Norway, between August 2016 and December 2017. Due to the spatial effects inherent in housing markets, the hedonic regression was enhanced with three widely utilized spatial econometric models and a fourth, outlier-robust model. This was done to ensure that any change in model performance was caused by methodological effects from the model combination, rather than being due to the correction of a spatially misspecified regression.

Historically, hedonic price regression models have been used when conducting house price analysis. First described by Rosen (1974) to value composite goods, this model is based on the assumption that the residential property value is merely the sum of the market value of its individual characteristics. Thus, accuracy of such hedonic house price predictions is determined by the data's ability to identify important housing attributes and ability correctly estimate the structural characteristics, time, and location as the main determinants of housing value. Although the first two factors require considerate specification, location modeling has proven particularly challenging in the classic hedonic regression framework. The issue primarily stems from the difficulty in capturing the spatial interactions in cross-sectional housing data, as these introduce simultaneity and feedback effects that necessitate use of spatial econometric models (Anselin, 2010). This has been a long-neglected fact in the studies of economics, arguably because spatial analysis is commonly associated with disciplines like geography and geology (Dubin, 1998).

The repeat sales model is another important real estate analysis methodology, based on the premise that the prices at which a particular property has been sold in the past are useful inputs for estimating future real estate market development (Bailey et al., 1963). When applying this model, it is common to multiply previous sale prices with the expected market growth to obtain current price estimates. While this method has been widely used, Case and Quigley (1991) demonstrated the merits of combining repeat sales with hedonic regression in the construction of house price indices. Their findings were subsequently confirmed and discussed by Case et al. (1991). However, extensive literature review has revealed that broader applicability of this combination has never been explored.[1]

To get robust results against spatial misspecification of the hedonic model, repeat sales price predictions are combined with the predictions yielded by regression models, which are further enhanced by applying both traditional and state-of-the-art spatial models reported in pertinent literature, as well as an *ad hoc* spatial model proposed in this paper. As a result, in addition to the primary contribution to the house price valuation methods, we provides further empirical evidence on the utility of spatial econometric modeling of the housing market, linking the analyses and findings reported in this paper to one of the most prevalent research trends in real estate valuation (Krause and Bitter, 2012).

Combining regression predictions with the estimates yielded by the repeat sales model improved the accuracy of all hedonic models with respect to all examined metrics. These improvements were attained even when novice combination techniques were adopted, which was primarily attributed to diversification effects (Bates and Granger, 1969). The geographically weighted regression outperformed the other spatial specifications, indicating that spatial non-stationarity is more prominent than spatial dependence in the Oslo housing market. Further, combining hedonic regression and repeat sales within a single model resulted in greater improvements to the outputs generated by regression models characterized by low accuracy. As the models differed in terms of location modeling only, it can be posited that repeat sales estimates contribute at least some spatial information to the overall model output. While the value of this contribution diminishes for well-specified spatial models, previous sales prices nonetheless likely contain a certain amount of non-spatial information that is otherwise difficult to discern from the market trends. If this assumption holds, previous sales prices could be particularly valuable for developing automated property valuation tools, as few alternatives for detecting such information exist besides human inspection.

The remainder of this article is organized as follows. First, the Norwegian housing market, and that characterizing Oslo in particular, is introduced in Section 2, while the data employed when testing the models is presented in Section 3. The real estate evaluation models are presented in Section 4, while their results are reported and discussed in Section 5. The main conclusions are presented in Section 6, along with some suggestions for future research directions in this field.

2 Background

2.1 The Norwegian property market

The Norwegian housing market has some noteworthy characteristics, making it highly suitable for studies on property pricing in general. First, the sales process can be characterized as an *English auction*, where the price is determined in a near perfect bidding context (Olaussen et al., 2017). Second, most properties for sale in Norway are announced via

[1] Extensive literature review has failed to uncover any publicly available research on this topic. However, some companies advertise automated valuation based on both models, e.g., Home Value Explorer® by Freddie Mac (2017).

standardized advertisements published on the FINN.no website.[2] Such high degree of transparency and standardization facilitates comparison between dwellings and provides high-quality data for market participants. Third, Norwegians have a strong preference for home ownership as opposed to renting, as indicated by the 82.7% ownership rate reported for 2016 by Eurostat (2016).

2.2 The property market of Oslo

Oslo is the capital of Norway with a 2018 population of approximately 670,000. Historically, the city has been demographically divided between east and west, whereby industry workers were based around the river Akerselva in the central and eastern areas, while wealthier families mainly resided in western parts (Amundsen, 2015). Even though some former working-class districts like Grünerløkka and Gamle Oslo are becoming increasingly popular (Faksvåg, 2015), the historical pattern with higher prices in western areas is still evident, as shown in Figure 3.1.

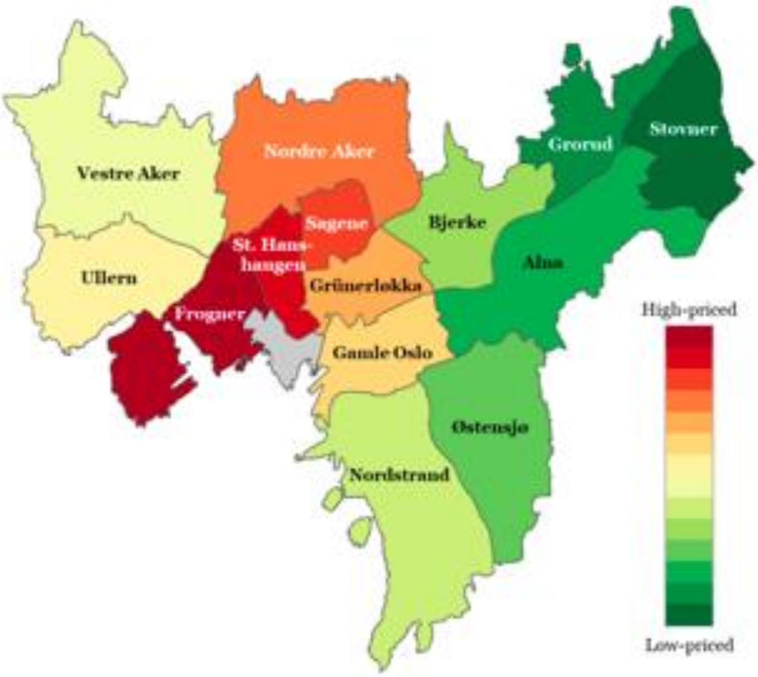


Figure 3.1: Administrative districts of Oslo with the price/m^2 ranking for 2017 given by Humberset (2018). Data for district Sentrum (denoted by grey color) were not available, while district Søndre Nordstrand is not represented in our dataset and is thus not shown on the map.

[2] FINN covers approximately 70% of the Norwegian housing market (Eiendom Norge, Eiendomsverdi and FINN.no, 2017). All properties in the dataset employed in the current investigation were announced on the site.

3 Data

The real property transaction data are compiled from the property register of Oslo, provided by the firm Alva Technologies (Alva). The dataset comprises of all housing transactions that took place in Oslo between August 2016 and December 2017. This dataset provides accurate and comprehensive information on building characteristics for each transaction, including longitude and latitude of the relevant residential property. Alva also provided the previous transaction prices for the dwellings if these were available. Prior to utilizing this data in the current investigation, some modifications were needed. For example, all entries related to the *Marka* district were discarded, dwellings from district *Sentrum* were reassigned to *St. Hanshaugen* and there was 36 dwellings labeled as *Other unit type*, after manual checking all 36 they was labeled as *Apartments*. This data preprocessing resulted in 16,417 residential units within the dataset, which was further augmented by mapping administrative district information from Oslo Kommune (2018), as well as by obtaining additional data by reviewing the corresponding FINN advertisements. An overview of the variables included in the regression models is provided in Table 3.1−3.5, where all attributes are specified as indicator variables. Variables derived directly from FINN are described in Table 3.5.

Parts of the information sourced from FINN were obtained through word recognition, which was applied to the advertisement title. As a result, only the property characteristics highlighted by the seller/agent were examined, potentially disregarding the attributes that a certain properties possess. However, the likelihood of missing potentially vital information was limited, as these titles are comprehensive, with the dwellings included in the dataset examined containing almost 17 words on average, which is sufficient for promoting multiple property characteristics. Further, since data related to variables that typically enhance property value, such as *Has a garden*, were also retrieved, the effect of aforementioned problem on the models presented here is negligible, as promoting such attributes is in the seller's interest.

In the repeat sales method, Statistics Norway's *Price index for existing dwellings for Oslo and Bærum* (Monsrud and Takle, 2018) was used as a proxy for the expected price appreciation, as it provides sales information from 1993 to the present. The distribution of the numbers of previous sales for the dwellings used in the repeat sales method is provided in Table 3.6.

Table 3.1: Construction year

Construction year	Dwellings
1820−1989	12,894
1990−2004	1,120
2005−2014	2,063
2015-2017	340

Total	16,417

Table 3.2: House type distribution

House type	Dwellings
Apartment	14,592
Semi-detached house	367
Detached house	385
Serial house	1,073
Total	16,417

Table 3.3 District distribution

District	Dwellings
Alna	1,317
Bjerke	719
Frogner	1,616
Gamle Oslo	1,647
Grorud	771
Grünerløkka	2,079
Nordre Aker	816
Nordstrand	1,038
Sagene	1,949
St. Hanshaugen	1,193
Stovner	556
Ullern	641
Vestre Aker	725
Østensjø	1,35
Total	16,417

Table 3.4: Size distribution

Dwelling Size	Dwellings
10—29 m^2	578
30—39 m^2	1,513
40—49 m^2	1,933
50—59 m^2	2,886
60—69 m^2	3,277
70—79 m^2	1,901
80—89 m^2	1,303
90—99 m^2	757
100—109 m^2	578
120—119 m^2	348
120—129 m^2	292

130—139 m^2	203
140—149 m^2	165
150—179 m^2	328
Above 180 m^2	355
Total	16,417

Table 3.5: Variables retrieved from FINN advertisements

	Dwellings	% of total
High monthly shared cost	1,642	10.0%
Two bedrooms & size < 60 m^2	932	5.7%
Three bedrooms & size < 85 m^2	835	5.1%
Housing cooperative	8,615	52.5%
Needs refurbishment	1,158	7.1%
Is a penthouse	2,299	18.2%
Has a garden	1,659	10.1%
Has a terrace	1,139	6.9%

High monthly shared cost is defined as being ranked within the top 10% for all dwellings in the dataset, the threshold being NOK 4,713 per month; Two bedrooms & size < 60 m^2 is the number of 2-bedroom dwellings covering area smaller than 60 m^2; Three bedrooms & size < 85 m^2 is the number of 3-bedroom dwellings with the floor area smaller than 85 m^2; Housing cooperative denotes whether the dwelling is part of a housing cooperative. The self-explanatory variables "Needs refurbishment," "Is a penthouse," "Has a garden" and "Has a terrace" were retrieved by word recognition.

Table 3.6: Number of sales for a dwelling in the dataset

Number of sales	Dwellings	% of total
One sales	3,279	20.0%
Two sale	5,631	34.3%
Three sales	4,109	25.0%
Four sales or more	3,398	20.7%
Total	16,417	100.0%

The table shows the number of dwellings and the share of the dataset where we have data on one sale, two sales, three sales and four or more sales.

4 Methodology

In this section, an ordinary hedonic regression model is introduced, with an emphasis on the intercept area dummy variables constructed using the k-means and k-nearest neighbor algorithms. Next, the following four extensions to the basic regression model are described: Regression kriging, Mixed regressive, spatial autoregressive model, Geographically weighted regression and Vicinity-based residual tuning.

Finally, the manner in which the estimates yielded by the repeat sales model are combined with the hedonic regression estimates is delineated. The schematic representation of the proposed models is given in Figure 5.1, where the geographically weighted regression is denoted by a dashed line, since variables related to districts must be omitted in this model.

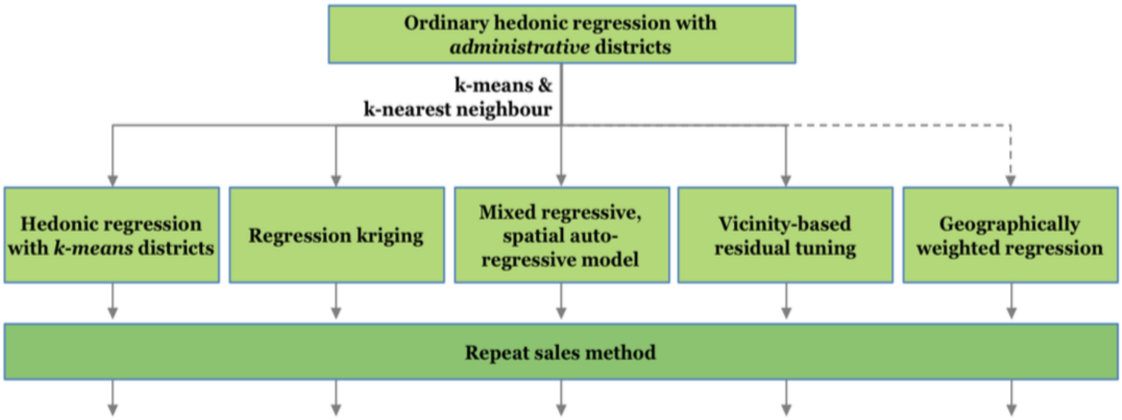


Figure 5.1: Overview of spatial models and extensions used in the present study. The dashed line indicates that district indicator variables cannot be specified in the GWR model.

4.1 Basic hedonic regression model

The hedonic regression model was first introduced by Rosen (1974). It has since been widely used in property valuation, due to the prevalent view that residential property value can be approximated by the sum of market value of its constituents. In the model employed, the value of a given dwelling is represented by the sum of its common debt[3] at sales and sales price, divided by the area in m^2, as given below:

$$P_i = \frac{sales\ price_i + common\ debt_i}{house\ area_i}. \quad (4.1)$$

The natural logarithm of P_i given by Equation (4.1) is estimated by evaluating contributions to the price by each utility-bearing attribute using multiple linear regression. The general equation to be estimated is given by:

$$\ln(P_i) = \beta_0 + \sum_k \beta_k X_{ki} + \sum_n \delta_n D_{ni} + \epsilon_i, \qquad \epsilon \sim i.i.d. \ (4.2)$$

where P is the price variable as defined in Equation (4.1); X_k is a set of explanatory variables, describing a presence of utility-bearing characteristic k (including both building characteristics and dummy variables pertaining to time); D_n is a set of n area indicator variables; ε is the error term, and β_0, β_k and δ_n are the parameters to be estimated, with their respective estimates denoted as $\hat{\beta}_0$, $\hat{\beta}_k$ and $\hat{\delta}_n$. As the data only span over 17 months, and cover a single city, the common assumption that parameter vectors are invariant across space and time is deemed valid (de Haan and Diewert, 2013). In line with the approach adopted by Koenker and Bassett Jr (1978), Equation (4.2) is estimated using *least absolute deviation*

[3] In Norway cooperatives and apartment buildings, can take on common debt for example to renovate the building. Especially for cooperatives, the common debt can be high compared with the transaction price of apartment. The total price of a dwelling in Norway is the transaction price plus the dwellings share of the total common debt.

(LAD), as LAD is more robust towards outliers than ordinary least squares (OLS) and other estimators based on distributional assumptions (Yoo, 2001). An overview of the structural explanatory variables used in Equation (4.2) is given in Table 3.1−3.5. To incorporate spatial and temporal variability in Equation (4.2), intercept indicator variables, further discussed in the following subsections, were introduced. Price predictions in nominal values is obtained by taking the exponential and multiplying with a scaling factor, to minimize underestimation bias in the transformation. The scaling-factor is estimated by regressing unscaled price estimates from the training sample on their corresponding real prices through the origin, where the 1% most expensive dwellings is discarded to control for outliers in the data.

House prices are volatile, as they are subject to seasonality and other effects, and are generally substantially influenced by time (Reichert, 1990). However, as the aim was modeling spatial effects, the temporal dimension is neglected. Specifically, to ensure that model output is unbiased by market price developments, this effect was isolated by including monthly time dummies as explanatory variables into all regression models. Furthermore, the test sample was constructed by randomly drawing 20% of the observations from the full sample. As a result of this approach, the two samples used for estimating and testing the models, respectively, span the same time horizon, thus eliminating the temporal dimension.

The k-means algorithm was applied to the training data sample to construct artificial market districts characterized by more homogeneous property pricing processes while retaining cohesiveness. For this purpose, distance between dwellings was measured as a function of longitude, latitude and price, whereby the latter was calculated by applying Equation (4.1). After clustering the training set, k-nearest neighbors were used to classify dwellings in the test sample based on the newly constructed districts, measuring distance in classical, geographic sense using the haversine formula (Sinnott (1984)). In empirical trials, the best results were obtained when the values of k in the k-means algorithm ranged from 14 to 20, and k = 18 was adopted in the final model, as this provided the most stable results. An illustrative plot comparing a k-means clustering with k = 14 and administrative borders is shown in Figure 6.1. The k in the k-nearest neighbor algorithm was set to 3, based on empirical trials, as well as visual inspection of district shapes produced by k-means.

4.2 Regression kriging

As argued by Dubin (1988) and Basu and Thibodeau (1998) among others, spatial dependence in the housing price process can be modeled by assuming that the original functional relationship given by Equation (4.2) holds, while abandoning the assumption of the error term being independent and identically distributed (i.i.d.), which requires modeling of the error covariance structure. Adopting this approach for prediction builds on the statistical interpolation technique known as kriging. Following the previously outlined notation, Equation (4.2) can be rewritten as:

$$\ln(P_i) = \beta_0 + \sum_k \beta_k X_{ki} + \sum_n \delta_n D_{ni} + \epsilon_i, \qquad \epsilon \sim N(0, \sigma^2 \mathrm{C}) \quad (4.3)$$

where C is the error correlation matrix. To estimate Equation (4.3), a functional form for the error term's covariance structure must be assumed. The parameters of this function, along with the normal regression coefficients, are simultaneously estimated using the maximum likelihood method. However, it should be noted that estimation of Equation (4.3) can become very complex when the dataset includes nonlinear explanatory variables (Hengl et al., 2007). Moreover, parameter instability is another major concern commonly encountered in practice (Goovaerts, 1999).

To mitigate these issues, the estimation process can be divided into two phases. First, the linear regression parameters β_0, β_k and δ_n are estimated using a less complex estimator—LAD, as previously noted. Next, the error covariance function parameters are estimated by *simple kriging*[4] with zero mean on the residuals from the first regression. The prediction process is finalized by adding the fitted residual from the simple kriging model to the fitted value from the linear regression. In mathematical terms, the predicted value for dwelling *i* in the test sample having structural characteristics X' and D' is given by

$$\ln(\hat{P}_i) = \hat{\beta}_{0,\mathrm{LAD}} + \sum_k \hat{\beta}_{k,LAD} X'_{ki} + \sum_n \hat{\delta}_{n,LAD} D'_{ni} + \sum_j w_{ij} \hat{\epsilon}_{j,LAD}, \ \ j \neq i \quad (4.4)$$

where $\hat{\varepsilon}$ are the LAD residuals from the training sample, and w_{ij} are the elements of the weight matrix W, determined by the *a priori* chosen covariance function. This two-step procedure was denoted as *regression kriging* by Odeh et al. (1995). Predictions yielded by Equation (4.4) and those resulting from directly estimating Equation (4.3) are mathematically equivalent. Indeed, Hengl et al. (2003) demonstrated that, as long as the assumed covariance function is identical, the difference is restricted to the computational steps.

Several structural covariance functions are applicable in kriging, provided that correlation between observations decreases with increased physical distance. In the current analyses, it is assumed that the error covariance follows the negative exponential form given below, as proposed by Case et al. (2004).

$$c_{ij} = b_1 + e^{-\frac{d_{ij}}{b_2}}, \qquad j = 1, 2, 3, \dots, 100; j \neq i$$

where the parameters b_1 and b_2 are estimated in the second step of the regression kriging procedure outlined above; d_{ij} are Euclidean distances between dwelling *i* and dwelling *j*, and c_{ij} are entries in the C-matrix derived from Equation (4.3). To calculate the weights based on the covariance matrix, the relationship $W = C^{-1}c$ was assumed, where c is a vector of covariances between the training data points and the estimation point (Bohling, 2005). To limit the computational cost, the number of neighbors for each dwelling was limited to 100.

At this juncture, it is important to note that generalized least squares (GLS) is typically recommended as the proper estimator in the first step of regression kriging, to account for spatial autocorrelation in the error term (Cressie, 1990). However, Kitanidis (1993) demonstrated that the difference between several iterations of GLS and a single iteration (OLS) is too small to have any notable effect on the final output. To test this claim, both GLS and LAD were adopted, yielding marginal differences, in line with the findings reported by Kitanidis (1993). Thus, LAD was chosen to ensure that a consistent choice of estimator is employed across all models evaluated.

[4] The term *simple kriging* is used when the mean of the dependent variable is assumed to be known *a priori* (Cressie, 1990).

4.3 Mixed regressive, spatial autoregressive model

As argued by Can (1992), spatial dependence in the housing price determination process can be modeled by including a function of the dependent variable as an autoregressive term in the standard hedonic regression (Equation (4.2)). Using the specification put forth by Fotheringham (2009, p. 257) the model can be expressed as:

$$\ln(P_i) = \beta_0 + \rho \sum_j w_{ij} \ln(P_j) + \sum_k \beta_k X_{ki} + \sum_n \delta_n D_{ni} + \epsilon_i, \quad j \neq i \quad (4.5)$$

where ρ is a measure of the overall level of spatial dependence among ($ln(P_i)$, $ln(P_j)$) pairs for which $w_{ij} > 0$, and w_{ij} are spatial weights assigned to the sales price of dwelling j. Other variables are as described in Section 3. Including the dependent variable to the right-hand side of the equation induces *simultaneity*; hence, estimating Equation (4.5) with OLS or LAD produces biased estimates. However, this approach is commonly adopted, as appropriate estimation using maximum likelihood is extremely challenging (Farber and Yeates, 2006). A different solution was proposed by Can and Megbolugbe (1997), who advocated for inclusion of an additional constraint, thus giving Equation (4.5) the following revised form:

$$\ln(P_{it}) = \beta_0 + \rho \sum_j w_{ij} \ln(P_{j,t-m}) + \sum_k \beta_k X_{ki} + \sum_n \delta_n D_{ni} + \epsilon_i, \quad m = 1,2,\ldots; \; j \neq i \quad (4.6)$$

The distinction between Equation (4.5) and (4.6) is that the dependent variables in the latter are determined at time t and are hence exogenous, rendering OLS and LAD unbiased estimators. The weighting function, again following Can and Megbolugbe (1997), is given by:

$$w_{ij} = \begin{cases} \dfrac{1/d_{ij}}{\sum_j (1/d_{ij})} & : d_{ij} < 1.5km \\ 0 & : d_{ij} \geq 1.5\ km \end{cases} \qquad j = 1,2,3,\ldots,15; \; j \neq i \ (4.7)$$

where d_{ij} are Euclidean distances between dwelling i and dwelling j, with j representing the 15 dwellings located closest to the dwelling i, and having earlier sales dates than dwelling i. In the special case where two dwellings share the same location, d_{ij} is set to 10 m, to ensure that Equation (4.7) is defined for all observations.

It is also worth noting that some dwellings in the dataset are situated at remote locations, and thus no relevant neighbors are available for defining the autoregressive term. The same issue arises for the oldest transactions within the sample. Thus, to retain the same number of observations for all models, the autoregressive term for the aforementioned dwellings was assumed to be equal to the average log(price) for the relevant district, with the price defined by Equation (4.1).

4.4 Geographically weighted regression

As argued by Wheeler and Calder (2007), the housing price process is non-stationary over space, the coefficients in the traditional hedonic regression represent the global "average" only. As a result, accurate predictions necessitate application of an enhanced regression model that permits parameter variation across space (Yao and Fotheringham, 2016). The geographically weighted regression method enables such a local parameter estimation. We adopt the notation given by Fotheringham et al. (2002, p. 52), resulting in a revised traditional regression framework given by:

$$\ln(P_i) = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) X_{ki} + \epsilon_i, \qquad (4.8)$$

where u_i and v_i denote the coordinates of the *i*th point in space, and $\beta_k(u_i, v_i)$ is a realization of the continuous function $\beta_k(u, v)$ at point *i*. Note that the location area indicator variable *D* from Equation (4.2) is omitted in Equation (4.8), which contains a greater number of unknown than observed variables. Consequently, at point *i*, Equation (4.8) is approximated by:

$$\ln(P_i) = \beta_0 + \sum_k \beta_k X_{ki} + \epsilon_i, \qquad (4.9)$$

The parameters β_0 and β_k are independently estimated for all *i* locations with dwellings in the test sample. Estimation is conducted by weighting the observations in accordance with their proximity to location *i*, and the parameters are chosen to minimize the weighted sum of squared residuals. In line with the approach proposed by Fotheringham et al. (2002), Equation (4.9) is estimated with the weights calculated using a Gaussian kernel function:

$$w_{ij} = e^{-0.5\frac{d_{ij}^2}{b}}, \qquad i \neq j \;\; (5.10)$$

where d_{ij} are the Euclidean distances between point *i* and *j*; *b* is denoted as bandwidth and is chosen by applying the cross-validation optimization approach described by Cleveland (1979). The practical implication of this choice is that only a small subset of the observations in the training sample is used to estimate Equation (4.9) at the different points *i*. Thus, the estimate for a given dwelling is vulnerable to anomalies in the data related to the neighboring dwellings.

4.5 Vicinity-based residual tuning

An automated variant—referred to as *vicinity-based residual tuning* or *VRT*—of a valuation method commonly used by real estate agents is adopted. In the original approach, a limited number of recently sold properties in the immediate neighborhood (usually 3−6) is used to provide a house value estimate (Can and Megbolugbe, 1997; Pace et al., 2000). The procedure outlined here is based on the premise that differences between properties are already controlled for in the residuals of a hedonic regression. Moreover, the issues that arise from including district intercept dummies in a regression, as outlined by Fik et al. (2003), are also addressed.

The fitted values for dwellings in the test set were obtained by using regression coefficients estimated on the training set. Next, for each dwelling in the test set, with the sales date denoted by τ. Up to κ closest neighbors from the training set sold before time τ, located within the same district[5] and within a radius of maximum μ meters were identified. The residuals of the neighbors were extracted before calculating their median, which was multiplied by a deflation factor α (along with another deflation factor β if the number of neighbors is below λ). Finally, this residual was added to the fitted value to obtain the VRT estimate, as shown in Table 4.1.

Table 4.1: Parameter values yielded by the VRT method

κ	μ	α	β	λ
6	150	0.7	0.5	3

Specifying area intercept dummies in a hedonic regression often results in low prediction accuracy close to district borders, where residuals with different magnitudes and signs are clustered on either side of the border. Figure A.2 provides an example of such effects. To address this issue, the district constraint was included in the step (i) above. Further, an outlier with an extreme residual value included as a neighbor can have a severe impact on the model accuracy. In the present investigation, this effect was mitigated by using the median and including the (λ, β) clause in the step (ii) above, where $\lambda = 3$ corresponds to the lowest number of neighbors where the smallest and largest neighbor residual value are discarded in the calculation of the median. The remaining model parameters were determined by applying the following reasoning: μ was chosen intuitively, and α and β values were determined by empirical trials, while the selection of κ was based on the approach recommended by Can and Megbolugbe (1997).

4.6 Constructing and combining repeat sales predictions

A common drawback of all hedonic house price models stems from the high heterogeneity among dwellings, rendering the inclusion of all price-influencing attributes infeasible (Case et al., 1991). To overcome this issue, repeat sales analysis was conducted, as this allowed capturing some of the effects that would be difficult to observe through the former sales prices of a given dwelling. The model is grounded in the assumption that residential property prices have developed in line with the overall market trends, as described by a house price index, implying that the quality of each dwelling is assumed comparable at the time of each sales transaction. As outlined in Section 3, Statistic Norway's *Price index for existing dwellings for Oslo and Bærum* provides quarterly data dating back to 1993. In adopting this resource, a maximum of three previous transactions for each dwelling was considered, giving preference to the most recent transactions, excluding all sales that occurred prior to 1993, as this period precedes the development of the aforementioned price index.

The premise that a dwelling's quality is highly similar at different transaction times is a questionable assumption. If previous sales conditions are unrepresentative for the dwelling's condition at resale, the repeat sales estimate is likely to be erroneous. To remove such outliers, all repeat sales estimates deviating by more than 25% from the regression estimate with which they are to be combined were discarded, in line with the approach recommended by Anon. (2013). To obtain one final prediction, the remaining estimates were combined

[5] Either administrative or generated by k-means, depending on the variable type required in the regression.

following a stepwise procedure. Briefly, the weight given to the hedonic regression estimate was at least 60%[6], and heavier weighting was given to predictions based on more recent previous sales than earlier transactions. In line with the approach utilized by Clemen (1989), only simple linear combination techniques were used.

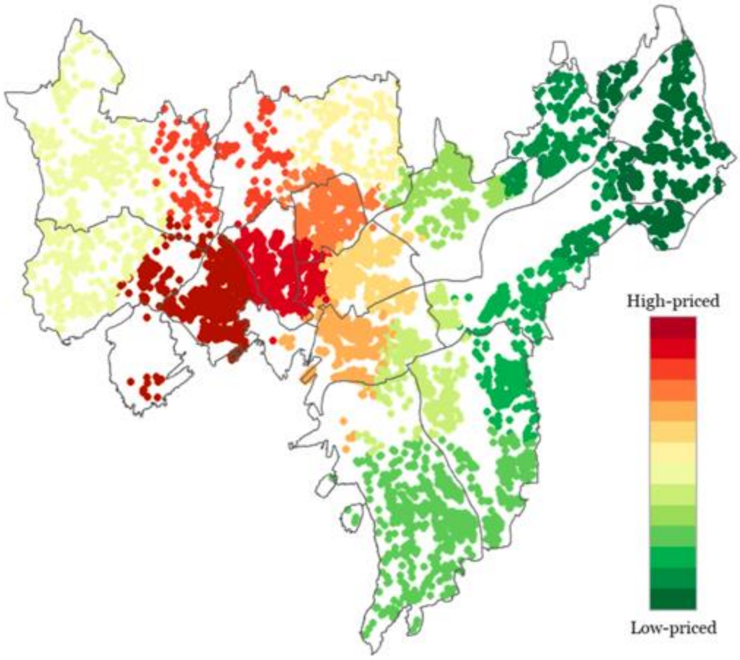
5 Results and discussion

The performance[7] of an ordinary hedonic regression without any location attributes is displayed in the top row of Table 5.1. As this model includes no spatial information, it represents a benchmark for assessing the utility of all enhancements incorporated into subsequent models in order to address the spatial aspect of residential property pricing. A comparison of the results confirms the strong influence of location on housing value. Indeed, addition of administrative district indicator variables only (row 2) reduces the median error from 12.1% to 8.05%, an improvement of 33.5%. Interestingly, augmenting the benchmark model with either regression kriging (row 5) or mixed regressive, spatial autoregressive model (row 9) yields similar improvements—from 12.1% to 8.18% and 7.70%, respectively. Thus, it can be argued that district intercept dummies incorporate the effect of location relatively accurately, although several methods can be adopted to address this issue. The extensive use of indicator variables is likely driven by the intuitive interpretation of the parameters, as well as ease of implementation. However, reliance on such variables, particularly when based on administrative districts, disregards intra-district variation and tends to result in irregular residual patterns close to borders. The resulting residual pattern from using administrative borders is plotted in Figure A.2.

Statistically generated districts can mitigate the aforementioned issues. A comparison of the administrative and a k-means based division of Oslo is depicted in Figure 5.1. As k-means operates independently of administrative districts, any area similarities are coincidental. An interesting case is the administrative district Alna, where k-means classifies the dwellings into four districts, indicating marked internal price differences. Further observations can be gleaned from comparing Figure 5.1 and Figures 3.1, as well as Figure A.1 and A.2 given in the Appendix. Improved performance from using k-means districts becomes evident when comparing row 2 and 3 in Table 5.1, as the median absolute percentage error improves from 8.05% to 7.67%. Moreover, the corresponding Moran's I and Geary's C values (Moran, 1950; Geary, 1954; Cliff and Ord, 1970) indicate reduced spatial autocorrelation in the residuals. As stated in Subsection 4.1.2, k-means is set to divide the city into a higher number of districts (18) than the administrative division (14), due to more stable performance. Results based on different values of k are shown in Appendix Table A.1, supporting the algorithm's conceptual advantages, as the improvement arising from implementing k-means with 14 districts is

[6] By testing different weighting and combinations we found no single optimal solution for multiple performance metrics - resulting in our choice of a "trail-and-error" based weight of 60% for the regression estimate that resulted on both high prediction and low volatility across multiple runs.

[7] Generally, model performance is measured by *median absolute percentage error* ($Q_{0.5}$).



relatively high compared to the improvement stemming from finer district fragmentation.

Figure 5.1: Comparison of administrative districts (lines) and statistically generated districts delineated by using k-means (colored markers) for the city of Oslo. Final k-means models are based on k = 18, whereas this map is obtained by using k = 14 for easier visual inspection of algorithm functioning. Color gradient indicates the average price/m^2 for each k-means district, and is thus directly comparable to the map provided in Figure 3.1. Number of observations is 13,133.

Table 5.1: Model performance

Model	Admin district	K-means	Repeat sales	$Q_{0.25}$	$\mathbf{Q_{0.5}}$	$Q_{0.75}$	Within 10 %	Moran's I	Geary's C	Row no.
Ordinary regression	☐	☐	☐	3.69%	**12.1**%	21.3%	42.4%	0.524	0.449	1
	☑	☐	☐	3.74%	**8.05**%	14.1%	59.6%	0.143	0.826	2
	☐	☑	☐	3.55%	**7.67**%	13.5%	61.6%	0.107	0.860	3
	☐	☑	☑	3.25%	**6.94**%	12.3%	66.0%	0.099	0.867	4
Regression kriging	☐	☐	☐	3.72%	**8.18**%	14.6%	58.5%	0.045	0.940	5
	☑	☐	☐	3.55%	**7.72**%	13.9%	61.1%	0.038	0.949	6
	☐	☑	☐	3.53%	**7.72**%	14.0%	61.2%	0.037	0.950	7
	☐	☑	☑	3.23%	**7.00**%	12.5%	65.3%	0.036	0.950	8
Auto-regressive	☐	☐	☐	3.52%	**7.70**%	13.7%	61.0%	0.086	0.889	9
	☑	☐	☐	3.40%	**7.31**%	13.0%	63.7%	0.068	0.902	10
	☐	☑	☐	3.39%	**7.28**%	12.9%	63.6%	0.065	0.906	11
	☐	☑	☑	3.18%	**6.77**%	11.8%	67.4%	0.063	0.906	12
VRT	☐	☐	☐	4.03%	**8.76**%	15.7%	55.6%	0.204	0.761	13
	☑	☐	☐	3.26%	**7.12**%	12.7%	64.6%	0.049	0.918	14
	☐	☑	☐	3.25%	**7.06**%	12.6%	64.8%	0.044	0.926	15
	☐	☑	☑	3.01%	**6.54**%	11.6%	68.2%	0.051	0.919	16
GWR			☐	3.14%	**6.65**%	11.7%	68.1%	0.072	0.896	17
			☑	2.93%	**6.20**%	11.0%	71.1%	0.059	0.909	18

Figure 5.2: Model refers to the methods outlined in Section 4.1−4.5; Admin district and K-means indicate if the boundaries for the area dummy variables are administrative districts or are generated by k-means, respectively (irrelevant for the GWR model); Repeat sales indicates whether the results are obtained after combining the output with the repeat sales predictions; $Q_{0.25}$, $Q_{0.5}$ and $Q_{0.75}$ denote the first, second and third error quartile, respectively, where $Q_{0.5}$ is boldfaced for emphasis; Within 10% specifies the fraction of errors below 10%; Row no. is row number provided for convenience when referring to this table. The results shown are average values based on the outputs of 10 runs for each implementation. The number of observations used for model training is 13,133, while the number of out-of-sample observations is 3,284.

District indicators are insufficient to appropriately model refined spatial patterns. Thus, the performance of the global augmentations—regression kriging and mixed regressive, spatial autoregressive model—is examined here first. Without incorporating district variables, both models display improved prediction accuracy compared to the benchmark model, as already pointed out. When district variables are included in the model, accuracy increases further, although not substantially. A less intuitive result is that the two spatial models seem indifferent to the choice of district representation, as indicated by a comparison of results reported in row 6 with those in row 7, as well as row 10 with row 11 in Figure 5.2, in contrast to the clear advantage of applying k-means to the ordinary regression. Two possible explanations can be offered for this finding. First, the influence of district dummy variables declines when location is concurrently modelled by several methods. This assertion is supported by a comparison of the absolute values of the location dummy parameters from the hedonic regressions with k-means districts and hedonic regressions with k-means districts and autoregressive term. Second, the two enhancements correct some spatial abnormalities caused by the administrative district, reducing the need for k-means. Finally, the autoregressive model outperforms regression kriging. However, this finding cannot be compared to previously published results, as none are available. Moreover, in line with LeSage and Pace (2014), applying different weighting functions in the regression kriging model did not affect the results. However, while this might reduce the credibility of the kriging implementation presented here, the effect of combining these predictions with repeat sales estimates (discussed later in this section) coincides with the remaining spatial models.

The VRT model performs second best among those aimed at spatial enhancements (as can be seen from the results reported in rows 14 and 15 in Figure 5.2). These findings are supported by the arguments put forth by Chan et al. (1999), who highlighted the severe impact of outliers on most models, which is avoided in VRT, as it is constructed to be more outlier-robust. It is also noteworthy that the VRT model only performs well for specifications including district variables,[8] likely due to the inability to distinguish more district-wide trends when considering a very limited number of neighbors. However, rather than adjusting the model to capture such trends, it is intrinsically tailored to address spatial residual patterns emerging from the use of intercept dummy variables in a regression. As a result, the method probably has limited use in general forecasting. Nonetheless, it is highly effective in this specific context. VRT also seems indifferent to the choice of district representation, most likely for the reasons suggested earlier for regression kriging and the autoregressive model.

The geographically weighted regression emerges as the most precise spatial enhancement (as shown in row 17 of Figure 5.2). Since this model assumes and addresses spatial non-stationarity, such significant improvement strongly suggests that this is the more prominent spatial effect in the Oslo housing market. The fact that GWR seems to outperform other spatial models for out-of-sample predictions corresponds with the findings reported by Farber and Yeates (2006) and Páez et al. (2008). However, it contrasts arguments put forth by Harris et al. (2010) and Harris et al. (2011), who recommended universal kriging. Although GWR tends to provide precise predictions, it has received criticism owing to its limited value for making inferences. Furthermore, the method is sensitive to outliers on a local level, which is

[8] Row 13 in Table 5.1 shows unsatisfactory performance by VRT where district variables are omitted.

particularly problematic in housing valuation, where outliers pose a permanent challenge.

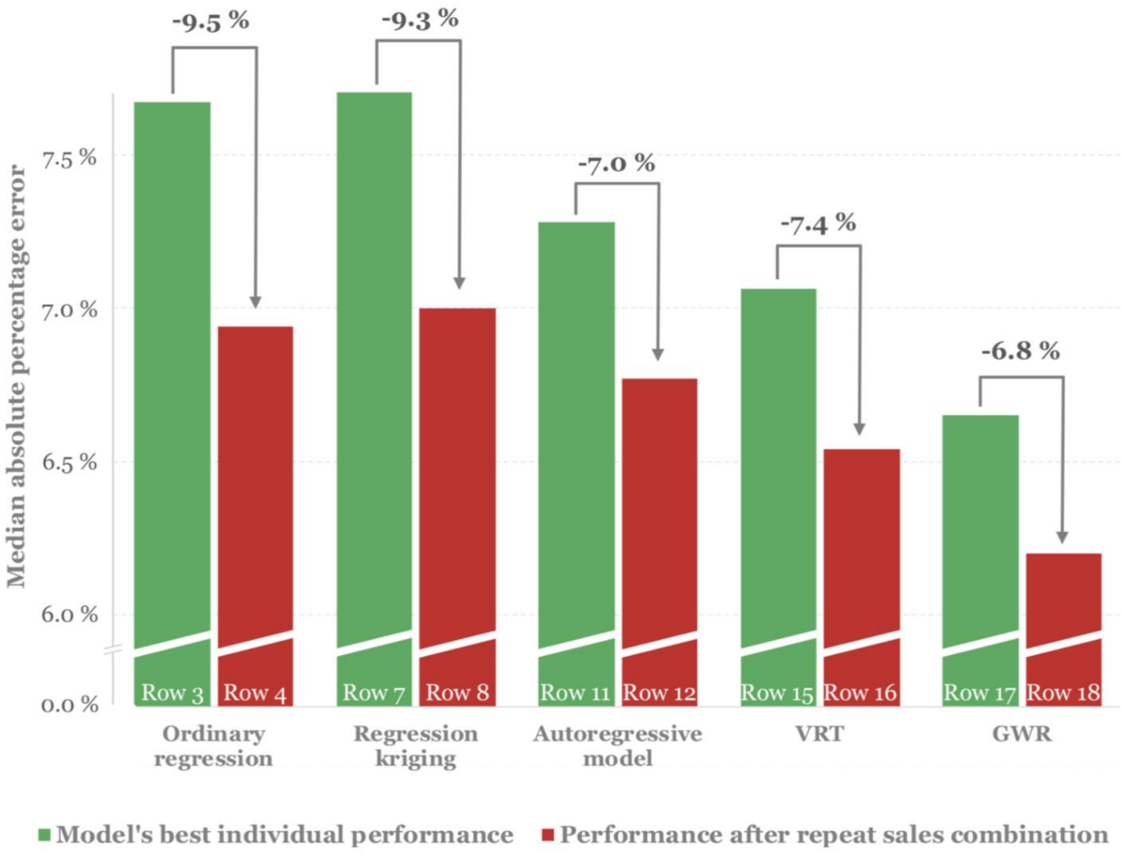


Figure 5.3: Visualization of improved performance achieved by combining repeat sales predictions with hedonic regression predictions. The bold number above the arrows indicates the reduction in median absolute percentage error in percentage terms. The row number at the bottom of each column indicates the corresponding row in Table 5.1.

The gain from combining repeat sales predictions with hedonic regression forecasts is evident in Figure 5.2 and is further emphasized in Figure 5.3, where the median absolute error achieved by the different regression models pre-and post-combination is plotted, which is equivalent to comparing rows 4, 8, 12, 16 and 18 with the corresponding values in Figure 5.2. In fact, the tabulated results reveal that combining repeat sales predictions with hedonic regression forecasts improves model accuracy by every metric and for every variation of the hedonic regression. To support diversification as the main driver behind this improvement, as argued by Bates and Granger (1969), as opposed to a deterioration of highly accurate repeat sales predictions, independent repeat sales results are provided in Appendix Table A.2. The data reported in this table confirm that, when used in isolation, repeat sales predictions are outperformed by all regression models incorporated into the combined models, supporting the diversification argument. It is also worth noting the considerable effect of outlier removal on the repeat sales estimates, which becomes evident when the two rows in Appendix Table A.2 are compared. This is arguably a necessity in order to replicate the level of improvement from the repeat sales/hedonic regression combination.

Apart from the overall increase in model accuracy, Figure 5.3 shows that improvements derived from combining repeat sales with other enhancements vary between the regression models, where a more substantial effect is observed when the initial regression error is large.

Considering the fact that the regression models only differ in terms of location modeling, it can be argued that repeat sales contribute at least some spatial information, the value of which diminishes for more sophisticated spatial models. Arguably, location is modeled fairly well in the autoregressive, VRT and GWR models, where combination with the repeat sales method resulted in similar improvements of 0.51, 0.52 and 0.45 percentage points, respectively. Consequently, it is reasonable to conclude that the predominant part of these improvements stems from the incorporation of non-spatial information omitted from the hedonic regression. Although it is not verifiable, this argument is supported by the inherent heterogeneity of dwellings, making inclusion of all price-influencing attributes in a regression framework infeasible (de Haan and Diewert, 2013).

Based on the preceding discussion, it can be posited that previous sales prices can provide specific value in two ways. Most importantly, they have the ability to incorporate information on difficult to observe attributes. This could have a pivotal value in automated property valuation, as there are few alternatives for detecting such information besides human inspection. Second, they enable the implementation of a scalable, parsimonious forecasting model, incorporating easily available attributes only, and relying on previous sales prices to incorporate information on the omitted, more market-specific attributes. As no universal hedonic specification presently exists (Bowen et al., 2001), local expertise remains necessary to identify relevant price-influencing attributes in a given market (Gelfand et al., 1998).

While we conceptual advantages of combining the hedonic regression and repeat sales methods are demonstrated, some practical limitations of such approach should also be noted. First, collecting previous sales price data reflecting current housing quality is generally hard, and can even be impossible in certain cases. Newly built dwellings obviously lack such data, but very old sales prices are not informative either, as they rarely represent the current state of the property (Case and Shiller, 1987). As a result, aforementioned combination might be less useful for markets where houses are traded less frequently, such as rural or suburban areas in which family homes predominate (Clapp et al., 1991). In addition, data will inevitably be lacking for some residential properties, preventing the method's applicability to all dwellings. Finally, the scale of the model improvement should also be considered when deciding if combining the modeling approaches is useful in practice. For example, in the present analyses, the median error of GWR was reduced from 6.65% to 6.20% (a 6.8% improvement) when combining the regression predictions with estimates from repeat sales. This rather marginal improvement might imply that the combination has little practical implication. Arguably, both models are good enough for obtaining an approximate value estimate. Nonetheless, neither is good enough to make end users confident in the results.

6 Conclusion

A central aspect of uncertainty in housing transactions is inaccurate property valuation. In this article, benefits derived by combining property price predictions yielded by two well-known valuation methods—repeat sales and hedonic regression—were investigated. The developed models were tested by applying them to 16,417 historical residential property transactions in Oslo, Norway. Due to the spatial effects inherent in housing markets, the hedonic regression was enhanced with three widely utilized spatial econometric models and a fourth, outlier-robust model. This was done to ensure that any change in model performance was caused by methodological effects from the model combination, rather than being due to the correction of a spatially misspecified regression.

The studied combination resulted in improved accuracy for all hedonic regressions on all metrics, which was attributed to diversification, as proposed by Bates and Granger (1969). Models with lower pre-combination accuracy yielded greater improvements, where reduction in median absolute percentage error ranged from 9.5% for the ordinary regression to 6.8% for the geographically weighted regression. This difference in the gains is argued to indicate that repeat sales predictions contribute at least some spatial information. While this contribution might have limited value for refined spatial models, the existence of some non-locational information in previous sales prices could nonetheless have pivotal value for automated property valuation, as there are few alternatives for detecting such information besides human inspection.

When interpreting the findings reported in this paper, certain limitations of the model combination should be noted. Specifically, non-existent or inapplicable previous sales price data in certain markets is inevitable. Optimizing the simple combination scheme presented in this paper is also a fruitful path for future studies, e.g., through more considerate implementation of the temporal dimension of previous sales. Similarly, improving repeat sales accuracy by, for example, applying local price indices, would be beneficial. Considering broader trends in automatic housing valuation, machine learning appears to be usurping the position as focal point of research at the expense of hedonic regression (Park and Bae, 2015). However, these tools remain highly dependent on the quality and quantity of observable, quantifiable data (Trawiński et al., 2017). Thus, given that previous sales prices seem to incorporate some otherwise difficult to capture information, a repeat sales/machine learning combination is an interesting direction for further research.

References

Anon. (2013) OECD, Eurostat, International Labour Organization, International Monetary Fund, The World Bank, and United Nations Economic Commission for Europe. Repeat sales methods. In *Handbook on Residential Property Prices Indices (RPPIs)*. OECD publishing, Paris: 67–72.

Amundsen, B. (2015). Rike og fattige flytter fra hverandre i Oslo [Rich and poor moveing apart from each other in Oslo]. *Forskning.no*. Available at: https://forskning.no/samfunnsgeografi/2015/06/rike-og-fattige-flytter -fra-hverandre-i-oslo (accessed: 11 December 2017).

Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, 89(1): 3–25.

Bailey, M. J., Muth, R. F., and Nourse, H. O. (1963). A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58(304): 933–942. Available at: http://www.jstor.org/stable/2283324 (accessed: 16 November 2017).

Basu, S. and Thibodeau, T. G. (1998). Analysis of spatial autocorrelation in house prices. *The Journal of Real Estate Finance and Economics*, 17(1): 61–85. Available at: https://link.springer.com/article/10.1023%2FA%3A1007703229507 (accessed: 1 May 2018).

Bates, J. M. and Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4): 451–468.

Bohling, G. (2005). Kriging. *Data Analysis in Engineering and Natural Science*. Kansas Geological Survey. Available at: http://people.ku.edu/~gbohling/cpe940/Kriging.pdf (accessed: 8 June 2018).

Bowen, W. M., Mikelbank, B. A., and Prestegaard, D. M. (2001). Theoretical and empirical considerations regarding space in hedonic housing price model applications. *Growth and Change*, 32(4): 466–490.

Brooks, C. (2014). *Introductory Econometrics for Finance*. Cambridge University Press, Cambridge, 3rd edition.

Can, A. (1992). Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics*, 22(3): 453–474.

Can, A. and Megbolugbe, I. (1997). Spatial dependence and house price index construction. *The Journal of Real Estate Finance and Economics*, 14(1-2): 203–222.

Case, B., Clapp, J., Dubin, R., and Rodriguez, M. (2004). Modeling spatial and temporal house price patterns: A comparison of four models. *The Journal of Real Estate Finance and Economics*, 29(2): 167–191.

Case, B., Pollakowski, H. O., and Wachter, S. M. (1991). On choosing among house price index methodologies. *Real Estate Economics*, 19(3): 286–307.

Case, B. and Quigley, J. M. (1991). The dynamics of real estate prices. *The Review of Economics and Statistics*, 73(1): 50–58.

Case, K. E. and Shiller, R. J. (1987). Prices of single-family homes since 1970: New indexes for four cities . NBER Working Paper Series 2393, National Bureau of Economic Research. doi:10.3386/w2393.

Chan, Y. L., Stock, J. H., and Watson, M. W. (1999). A dynamic factor model framework for forecast combination. *Spanish Economic Review*, 1(2): 91–121.

Chen, Y. (2016). Spatial autocorrelation approaches to testing residuals from least squares regression. *PloS one*, 11(1): e0146865.

Clapp, J. M., Giaccotto, C., and Tirtiroglu, D. (1991). Housing price indices based on all transactions compared to repeat subsamples. *Real Estate Economics*, 19(3): 270–285.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4): 559–583.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368): 829–836.

Cliff, A. D. and Ord, K. (1970). Spatial autocorrelation: a review of existing and new measures with applications. *Economic Geography*, 46: 269–292.

Corcoran, C. and Liu, F. (2014). Accuracy of Zillow's home value estimates. *Real Estate Issues*, 39(1): 45–49.

Cressie, N. (1990). The origins of kriging. *Mathematical Geology*, 22(3): 239–252.

de Haan, J. and E. Diewert (2013), "Hedonic Regression Methods", in *Handbook on Residential Property Price Indices*, OECD Publishing,
Paris, doi.org/10.1787/9789264197183-7-en.

Dubin, R. A. (1988). Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *The Review of Economics and Statistics*, 70(3): 466–474.

Dubin, R. A. (1998). Spatial autocorrelation: a primer. *Journal of Housing Economics*, 7(4): 304–327.

Eiendom Norge, Eiendomsverdi and FINN.no. (2017). Eiendom Norges boligprisstatistikk [Real Estate Norway housing price statistics]. Available at:
http://eiendomnorge.no/wpcontent/uploads/2017/11/Boligstatistikk_oktober_01.pdf
(accessed: 4 December 2017).

Eurostat. (2016). Distribution of population by tenure status, type of household and income group − EU-SILC survey. Available at:
http://appsso.eurostat.ec.europa.eu/nui/show.do?datasetÆilc_lvho02&langÆen (accessed: 20 May 2018).

Farber, S. and Yeates, M. (2006). A comparison of localized regression models in a hedonic house price context. *Canadian Journal of Regional Science*, 29(3): 405–420.

Fik, T. J., Ling, D. C., and Mulligan, G. F. (2003). Modeling spatial variation in housing prices: a variable interaction approach. *Real Estate Economics*, 31(4): 623–646.

Fotheringham, A. S. (2009). Geographically weighted regression. In S. Fotheringham & P. Rogerson (Eds.) *The SAGE Handbook of Spatial Analysis*, pages 243–254. SAGE Publications Ltd, London.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, New York.

FreddieMac. (2017). Home Value Explorer. Available at: http://www.freddiemac.com/hve/hve.html (accessed: 5 December 2017).

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5(3): 115–146.

Gelfand, A. E., Ghosh, S. K., Knight, J. R., and Sirmans, C. F. (1998). Spatio-temporal modeling of residential sales data. *Journal of Business & Economic Statistics*, 16(3): 312–321.

Goovaerts, P. (1999). Using elevation to aid the geostatistical mapping of rainfall erosivity. *Catena*, 34(3-4): 227–242.

Harris, P., Brunsdon, C., and Fotheringham, A. S. (2011). Links, comparisons and extensions of the geographically weighted regression model when used as a spatial predictor. *Stochastic Environmental Research and Risk Assessment*, 25(2): 123–138.

Harris, P., Fotheringham, A., Crespo, R., and Charlton, M. (2010). The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. *Mathematical Geosciences*, 42(6): 657–680.

Helbich, M., Brunauer, W., Hagenauer, J., and Leitner, M. (2013). Data-driven regionalization of housing markets. *Annals of the Association of American Geographers*, 103(4): 871–889.

Hengl, T., Heuvelink, G. B. M., and Rossiter, D. G. (2007). About regression-kriging: from equations to case studies. *Computers & Geosciences*, 33(10): 1301–1315.

Hengl, T., Heuvelink, G. B. M., and Stein, A. (2003). Comparison of kriging with external drift and regression kriging. Technical note, ITC. Available at: https://webapps.itc.utwente.nl/librarywww/papers_2003/misca/hengl_comparison.pdf (accessed: 11 May 2018).

Humberset, K. (2018). Her må du punge ut 85.000 kroner for én kvadratmeter [Here you have to pay 85,000 NOK for one square meter]. *aftenposten. no*. Available at: https://www.aftenposten.no/bolig/Her-ma-du-punge-ut-

85000-kroner-for-n-kvadratmeter-10982b.html (accessed: 4 June 2018).

Kitanidis, P. K. (1993). Generalized covariance functions in estimation. *Mathematical Geology*, 25(5): 525–540.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 46(1): 33–50.

Krause, A. L. and Bitter, C. (2012). Spatial econometrics, land values and sustainability: Trends in real estate valuation research. *Cities*, 29: S19–S25.

LeSage, J. P. and Pace, R. K. (2014). The biggest myth in spatial econometrics. *Econometrics*, 2(4): 217–249.

Levin, J. (2001). Information and the market for lemons. *RAND Journal of Economics*, 32(4): 657–666. Available at: http://www.jstor.org/stable/2696386 (accessed: 27 May 2018).

Monsrud, I. J. and Takle, M. (2018). Price index for existing dwellings. *Statistics Norway*. Available at: https://www.ssb.no/en/priser-og-prisindekser/statistikker/bpi (accessed: 7 June 2018).

Moran, P. A. (1950). A test for the serial independence of residuals. *Biometrika*, 37(1/2): 178–181.

Odeh, I. O. A., McBratney, A. B., and Chittleborough, D. J. (1995). Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression kriging. *Geoderma*, 67(3-4): 215–226.

Olaussen, J. O., Oust, A., and Solstad, J. T. (2017). Energy performance certificates–Informing the informed or the indifferent? *Energy Policy*, 111: 246–254.

Oslo Kommune. (2018). Finn bydelen din her [Find your city district here]. Available at: https://www.oslo.kommune.no/politikk-og-administrasjon/bydeler/bydelsvelger/ (accessed: 13 May 2018).

Pace, R. K., Barry, R., Gilley, O. W., and Sirmans, C. F. (2000). A method for spatial–temporal forecasting with an application to real estate prices. *International Journal of Forecasting*, 16(2): 229–246.

Páez, A., Long, F., and Farber, S. (2008). Moving window approaches for hedonic price estimation: an empirical comparison of modelling techniques. *Urban Studies*, 45(8): 1565–1581.

Park, B. and Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Systems with Applications*, 42(6): 2928–2934.

Reichert, A. K. (1990). The impact of interest rates, income, and employment upon regional housing prices. *The Journal of Real Estate Finance and Economics*, 3(4): 373–391.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82(1): 34–55.

Sinnott, R.W. (1984). Virtues of the haversine. *Sky and Telescope*, 68(2): 159.

Statistics Norway, House Price Index, Available at: https://www.ssb.no/statbank/table/07221/ (accessed: 5 February 2018).

Trawiński B., Lasota T., Kempa O., Telec Z., Kutrzyński M. (2017) Comparison of Ensemble Learning Models with Expert Algorithms Designed for a Property Valuation System. In: Nguyen N., Papadopoulos G., Jędrzejowicz P., Trawiński B., Vossen G. (eds) *Computational Collective Intelligence*. ICCCI 2017. Lecture Notes in Computer Science, vol 10448. Springer, Cham.

Wheeler, D. C. and Calder, C. A. (2007). An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems*, 9(2):145–166.

Yao, J. and Fotheringham, A. S. (2016). Local spatiotemporal modeling of house prices: a mixed model approach. *The Professional Geographer*, 68(2): 189–201.

Yoo, S.-H. (2001). A robust estimation of hedonic price models: least absolute deviations estimation. *Applied Economics Letters*, 8(1): 55–58.

Appendix

Table A.1: K-means results with varying k, compared to administrative districts

District type	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	Within 10%	Moran's I	Geary's C
Administrative	3.74%	8.05%	14.1%	59.6%	0.143	0.826
K-means (k = 14)	3.61%	7.78%	13.9%	60.8%	0.125	0.839
K-means (k = 16)	3.66%	7.78%	13.9%	60.8%	0.115	0.853
K-means (k = 18)	3.55%	7.67%	13.5%	61.6%	0.107	0.860
K-means (k = 20)	3.54%	7.62%	13.6%	61.5%	0.098	0.870
K-means (k = 22)	3.57%	7.64%	13.6%	61.5%	0.103	0.866

District type refers to the type of district indicator used; $Q_{0.25}$, $Q_{0.5}$ and $Q_{0.75}$ denote the first, second and third error quartile, respectively. Within 10% specifies the fraction of errors below 10%, and values for Moran's I and Geary's C is presented. The tabulated results pertain to 14 administrative districts and represent average values from 10 runs for each implementation. The number of observations used for model training is 13,133, while there were 3,284 out-of-sample observations.

Table A.2: Independent repeat sales results

Model	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	Within 10%	Number of Observations	Percentage of Total
Repeat sales	4.11%	8.88%	16.1%	54.8%	13,138	80.0%

Repeat sales after outlier removal	3.68%	7.77%	13.4%	61.5%	11,69	71.2%

Model refers to whether the repeat sales estimates have undergone outlier removal; Q0.25, Q0.5 and Q0.75 denote the first, second and third error quartile, respectively and Within 10% specifies the fraction of errors below 10%. Number of Observations denotes the number of entries with previous sales prices, and Percentage of Total indicates the latter number as a percentage of the entire data set. The similarity between the number of entries with previous sales before outlier removal (13,138) and the number of observations used for model training (13,133) is purely coincidental.

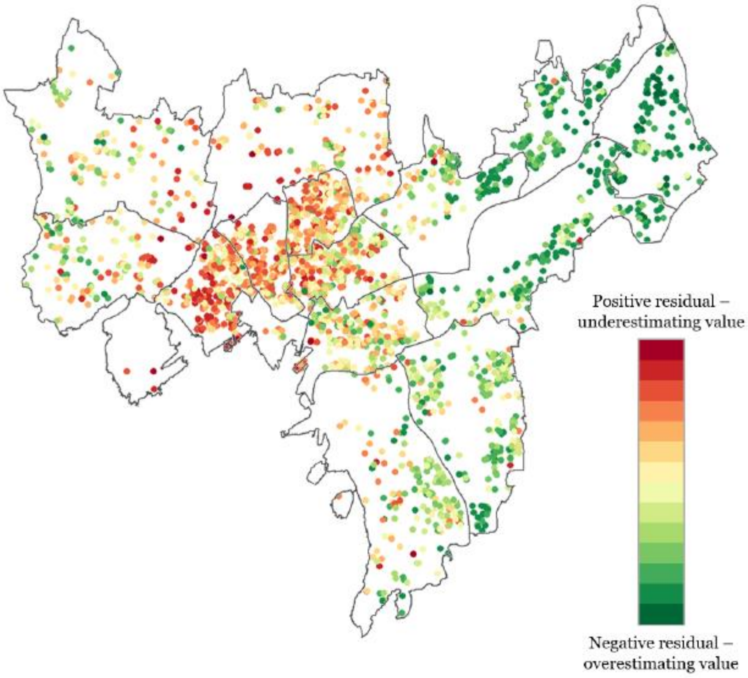


Figure A.1: Out-of-sample residuals from the hedonic regression without spatial enhancements or district indicators (referred to in Section 5the main text as the benchmark model) depicted on the map of Oslo. Lines represent administrative district boundaries, while marker color indicates residual value for each dwelling. Observations: 3,284.

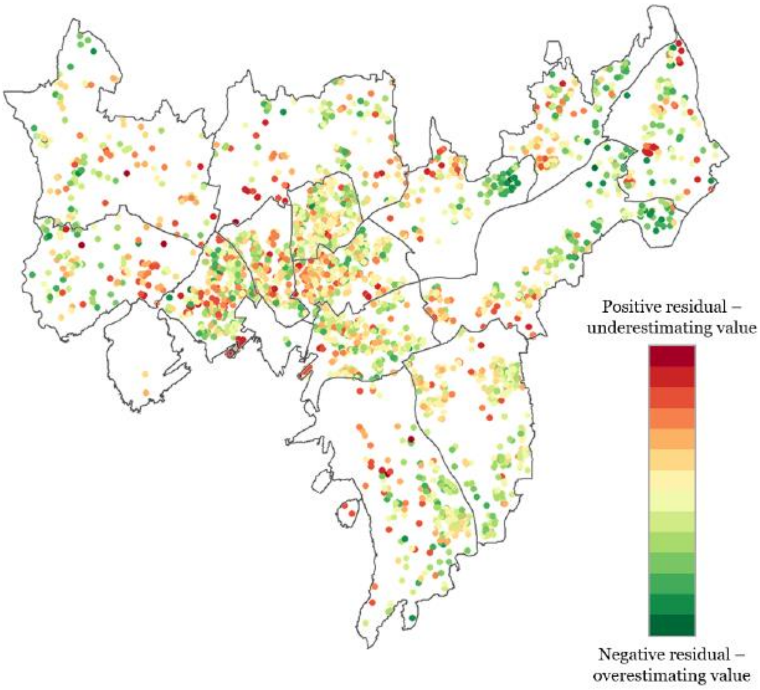


Figure A.2: Out-of-sample residuals from the hedonic regression with administrative district indicators depicted on the map of Oslo. Lines represent administrative district boundaries, whereas marker color indicates residual value for each dwelling. Observations: 3,284.

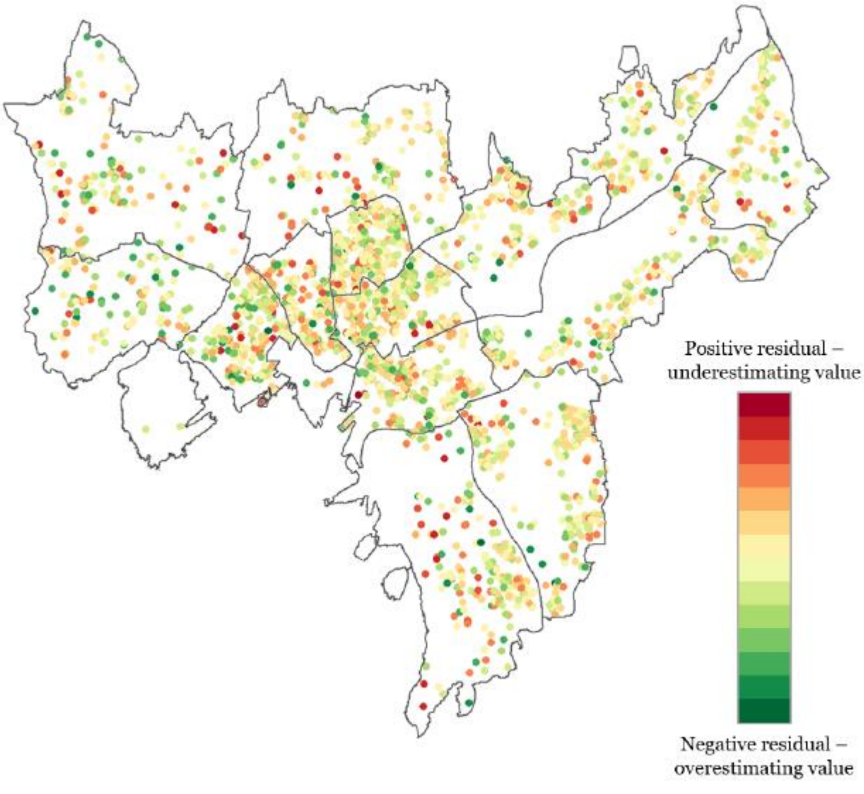


Figure A.3: Out-of-sample residuals from the most accurate model that combines geographically weighted regression and repeat sales depicted on the map of Oslo. Lines represent administrative district boundaries, whereas marker color denotes residual value for each dwelling. Observations: 3,284.