

Improved Methods for Predicting Property Prices in Hazard Prone Dynamic Markets

Koen de Koning¹  · Tatiana Filatova¹ · Okmyung Bin²

Accepted: 15 October 2016 / Published online: 26 October 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Property prices are affected by changing market conditions, incomes and preferences of people. Price trends in natural hazard zones may shift significantly and abruptly after a disaster signalling structural systemic changes in property markets. It challenges accurate market assessments of property prices and capital at risk after major disasters. A rigorous prediction of property prices in this case should ideally be done based only on the most recent sales, which are likely to form a rather small dataset. Hedonic analysis has been long used to understand how various factors contribute to the housing price formation. Yet, the robustness of its assessment is undermined when the analysis needs to be performed on relatively small samples. The purpose of this study is to suggest a model that can be widely applicable and quickly calibrated in a changing environment. We systematically study four statistical models: starting from a typical standard hedonic function and gradually changing its functional specification by reducing the hedonic analysis to some basic property characteristics and applying kriging to control for neighbourhood effects. Across different sample sizes we find that the latter performs consistently better in the out-of-sample predictions than other traditional price prediction methods. We present the specific improvements to the traditional spatial hedonic model that enhance the model's prediction accuracy. The improved model can be used to monitor price changes in risk-prone areas, accounting for changes in flood risk and at the same time controlling for autonomous market responses to flood risk.

✉ Koen de Koning
k.dekoning@utwente.nl

Tatiana Filatova
t.filatova@utwente.nl

Okmyung Bin
BINO@ecu.edu

¹ Department of Governance and Technology for Sustainability (CSTM), University of Twente, P.O. Box 217, 7500AE Enschede, The Netherlands

² Department of Economics, Center for Natural Hazards Research, East Carolina University, Greenville, NC 27858, USA

Keywords Hedonic analysis · Price prediction · Flood risk · Natural hazards · Kriging · Climate change · Small sample · Out-of-sample prediction

1 Introduction

Housing contributes largely to the welfare of individuals. Consequently, housing prices can strongly influence households' financial decisions (Bostic et al. 2005) making households richer or poorer as prices fluctuate. Changes in property prices are driven by changes in macro-economic conditions, changes in consumer preferences and incomes, and exogenous shocks (Filatova 2014). At the times of natural disasters price tends to shift significantly and abruptly (Bin and Landry 2013; Atreya 2013) implying that there are systemic changes in property markets. In other words, transactions in the past may not be representative anymore when making current price assessments or projections for the future. Therefore it becomes important to utilize most recent sales in conducting reasonable market price assessments or predictions. Various comprehensive methods have been developed for these purposes in the past decades (Basu and Thibodeau 1998; Case et al. 2004; Dubin 1999; Pagourtzi 2003). Real estate appraisers, local taxation offices, mortgage lenders and insurance companies are eager to know the current value of properties in line with changing market conditions. Models that predict housing values should, thus, be calibrated with most recent sales that represent current developments in the market. There is much demand for models that can detect and predict trends in the market at an early stage, and they require robust predictions while being calibrated with only few observations (Kuntz and Helbich 2014). This is generally problematic in hedonic analysis that may require thousands of transactions to deliver reliable statistically significant estimates for various structural and spatial attributes that influence housing prices in a particular market.

Assessment of a value at risk is also an important part of cost-benefit analyses (CBA) in the context of natural hazards and risk mitigation policies. Valuation of capital at risk is an essential part of the direct damage estimate in any CBA and provides a tool for policy makers to efficiently allocate resources among competing risk management options. Flood risk is one of the most frequently occurring disasters worldwide, and CBA is widely applied to assess flood management strategies (Gamper et al. 2006; Hall et al. 2005; Merz 2010; Penning-Rowsell et al. 2005; Hallegatte 2006). Usually CBA's for flood risk rely on combining geographic (GIS) maps with flood zones (with probabilities and potential inundation depths), damage functions and land use data (Dutta et al. 2003; Hall et al. 2003, 2005). Flood risk is the sum of total impacts and probabilities of flood events with a particular severity and inundation depth:

$$\text{Flood risk} = \sum_{i=1}^{i_{\max}} P(X_i) * D(X_i) * K \quad (1)$$

where X is a list of all possible flood scenarios, $P(X)$ a list of all related probabilities, $D(X)$ is the damage to a property as a function of inundation depth, water stream speed and salinity (often expressed as a percentage of a property destroyed), and K is the market value of properties located within the flood zone. Given the growing concerns for increasing vulnerability of urban areas driven by climate change and a need for climate adaptation policies, a majority of the studies focus either on calculating new probabilities ($P(X)$), Eq. 1 (Hirabayashi 2013; Ward et al. 2014) or on estimating damage functions ($D(X)$), Eq. 1 for properties and infrastructure (Farber 1987; Oliveri and Santoro 2000; Merz 2010). Thus, while a lot of attention goes to clarifying location-specific hazard probabilities and

relations between severity of hazards and corresponding damages to properties, the value of capital at risk is assumed to remain static. Possible structural changes in property markets driven by, for example, increasing severity and frequency of flooding, are not considered. This approach is insufficient in a changing environment, especially when climate-related natural hazards are concerned. While little attention is currently given to changes in capital in hazard zones (K , Eq. 1), several hedonic studies documented that flood risk premium is not stable over time. Namely, values of flood-prone properties drop significantly after a flood event, but recover back just after a few years (Atreya et al. 2012; Bin and Landry 2013; Pryce et al. 2011). It appears that recent experience with flooding awakens or reinforces the perceived risks and costs associated with flooding, and that a lack of flooding experience vanishes these perceptions. Thus, flood risk assessments in CBA may be quite sensitive to the timing when a flood discount is measured or to the year of a property valuation. It is important to keep track of these market responses to floods driven by exogenous shocks and changes in individual risk perceptions and location choices, and to update the expected prices and the corresponding value of the capital at stake.

Hedonic analysis is commonly used to assess and predict property prices and to estimate the flood risk premiums. In hedonic analysis a list of housing attributes is combined into a multiple regression with sales price as dependent variable. It can be used to predict future sales prices, yet the main purpose of these models is to calculate the marginal implicit price of specific housing attributes such as neighbourhood amenities, environmental quality or safety against floods (Atreya et al. 2012; Bin and Polasky 2004; Bin and Landry 2013; Hallstrom and Smith 2005). Hedonic studies often employ a large scale cross-sectional data measured within a long time frame. The question remains whether these models can effectively predict prices when calibrated with only few recent sales. One of the problems with assessing and predicting future sales prices using traditional hedonic models, is the chosen functional relation between spatial factors and sales prices. The fact that housing location has a strong effect on sales price is widely acknowledged, but the complexity of space as a factor is not captured well enough in the hedonic literature (Dubin 1992). There are several ways to construct regression models that account for spatial and neighbourhood characteristics, including for example spatial error model (Anselin 2001). An extensive analysis of out-of-sample prediction performance of various spatial (econometric) models has been performed by Voltz and Webster (1990), Bourassa et al. (2007) and Basu and Thibodeau (1998). While usually hedonic analysis (including spatial error models) performs well on large multi-year datasets, there is a need for an improved approach for robust assessment of property prices in highly dynamic markets. As discussed above, dramatic price changes in property market suffering from a shock, such as flooding for example, require a price prediction model that can work on small samples such as a few months of transaction data.

To mitigate the problem of a careful and robust assessment of the influence of spatial factors, Dubin (1992) suggests to omit all spatial variables in the hedonic analysis and to interpolate the spatial correlation in property prices by using kriging. Kriging is a spatial statistics method used to perform spatial interpolation, and is used for a wide range of applications in environmental sciences also based just on few observation points (Alemi et al. 1988; Delhomme 1978; Hernandez-Stefanoni and Ponce-Hernandez 2006; Webster and Burgess 1983). Yet just a few hedonic studies have adopted this method despite the fact that it can significantly improve the prediction performance compared to the traditional regression-based hedonic analysis (Case et al. 2004; Kuntz and Helbich 2014). Some studies applied the technique to correct for spatial autocorrelation (Basu and Thibodeau 1998; Bourassa et al. 2007; Militino et al. 2004), and other studies also validated the method through out-of-sample predictions (Case et al. 2004; Kuntz and Helbich 2014). The model specifications examined

in this study are based on hedonic analysis and kriging. While the literature suggests that kriging improves the prediction performance of spatial hedonic models, the performance of these models over a range sample sizes have yet to be tested. Therefore, the main purpose of this paper is to assess the robustness of the prediction performance of spatial hedonic models, either enhanced or not with kriging, under different sample sizes.

Another method used to predict property prices is artificial neural networks (Nguyen and Cipps 2001). While housing attributes in hedonic models are typically fitted with linear, log or squared relationships with price, artificial neural networks are used to fit more complex functional relationships. This works well with large housing transactions samples, but is sensitive to over-fitting when calibrated with small samples. Nguyen and Cipps (2001) have compared the performance of multiple regression models with artificial neural network models across sample sizes, and have concluded that the multiple regression models perform better than artificial neural network models at small sample sizes. Moreover, regression models are far less complicated and more widely used than artificial neural networks, thus we do not consider it further in our paper.

Given the number of different methods to assess and predict property prices, the purpose of this study is to understand which model can be widely applicable across a range of sample sizes and can be quickly calibrated in a changing environment. Our main research objective is to determine which specification of a spatial hedonic model is the best in predicting sales prices when calibrated with a small sample of recent sales. We test various hedonic models by systematically changing the size of the in-sample set of transactions based on which the models are calibrated. The sales prices of out-of-sample properties are predicted with the calibrated models. We perform this analysis on the dataset with residential property transactions between 1992 and 2002 in a housing market in North Carolina. We also analyse the reliability of the flood risk discount assessed with different statistical models under various in-sample sizes. Given the challenges of assessing capital at risk and flood risk discount in particular in a changing environment, the outcomes of the current paper may be of interest for policy makers conducting CBA of flood risk management policies, for monitoring developments in insured and uninsured property values and capital-at-risk, and for assessing structural changes in property markets in response to natural hazards. The analysis in this paper can be applied to a wide range of natural hazards and real estate appraisal in general. Our results demonstrate when and why the kriging-enhanced hedonic model performs better. Especially the prediction performance with small sample sizes is interesting, because this is where the model can quickly be calibrated and be applied for price predictions under changing market conditions. We present the specific improvements to the traditional spatial hedonic model that enhance the prediction accuracy, especially when it is calibrated with few observations.

The paper proceeds as follows. We start by giving a description of the data and the four different models, which are systematically compared for different samples sizes. Then, we explain how the analysis is done to compare the models. We conclude by discussing the results and their implications.

2 Methods

2.1 Data

We use housing sales data in Pitt County, North Carolina, from January 1992 to June 2002 (Bin and Landry 2013) to calibrate the models and to validate their prediction performance.¹

¹ Note, that the original study of Bin and Landry (2013) employed a spatial error model. Kriging and spatial error models differ in the way in which the spatial weight matrix is constructed. The spatial weight matrices

Table 1 Summary statistics of the property attributes

Variables	Summary (N = 4779)	
	Mean	SD
Sales price	USD 156,612	87,354
Age of the house	22.4 years	19
Number of bedrooms	3.2	0.59
Total structure square feet	2391	993
Lot size in acres	0.64	2.4
Gas heating (= 1)	0.35	0.48
Fireplace (= 1)	0.77	0.42
Face brick (= 1)	0.48	0.50
Hard wood floor (= 1)	0.25	0.43
Good quality (= 1)	0.031	0.17
Vacant home (= 1)	0.0048	0.069
Distance to creek	854 feet	596
Distance to airport	33,966 feet	17,859
Distance to major road	135 feet	99
Distance to business centre	4632 feet	2452
Distance to railroad	5498 feet	6378
Distance to Tar River	20,999 feet	17,587
Distance to park	7490 feet	7051
Sold between Fran and Floyd (= 1)	0.34	0.48
Sold after Floyd (= 1)	0.37	0.47
Floodplain (= 1)	0.064	0.24

The area provides an excellent natural experiment setting for this study in that it had enjoyed a period of relative calm, not experiencing major hurricane flooding since Hurricane Hazel in 1954, followed by two major hurricanes. Hurricane Fran (1996) produced millions of dollars in property damages resulting from profuse rainfall, flash floods, and severe storm surge. Three years later Hurricane Floyd (1999) brought torrential rains and record flooding which resulted in one of the largest peacetime evacuations in U.S. history (Bin and Polasky 2004). The data include information on sales price, property characteristics such as age and size, dummy variables that represent the presence or absence of extra facilities of the property, spatial information on the distance to amenities and disamenities, flood zoning, and time of sales. The summary statistics of all the relevant property characteristics can be found in Table 1.

2.2 Model Specifications

The hedonic price function of a property is given by

Footnote 1 continued

in spatial error models are constructed based on the assumptions of the user, whereas in kriging they are based on the spatial structure of the error, which is defined in the construction of the semivariogram. The spatial error model is particularly relevant for proper estimation of the coefficients in the hedonic price estimation, whereas kriging focuses on the prediction of the dependent variable.

Table 2 Input variables and their functional forms for the hedonic models

Variables	M1	M2	M3	M4
Age of the house	$X + X^2$	$X + \sqrt{X}$	X	X
Number of bedrooms	$X + X^2$	$X + \sqrt{X}$	X	X
Lot size in acres	$X + X^2$	$X + \sqrt{X}$	\sqrt{X}	\sqrt{X}
Total structure square feet	$X + X^2$	$X + \ln(X)$	$\ln(X)$	$\ln(X)$
Gas heating (=1)	X	X		
Fireplace (=1)	X	X		
Face brick (=1)	X	X		
Hard wood floor (=1)	X	X		
Good quality (=1)	X	X		
Vacant home (=1)	X	X		
Log of distance to creek	X	X		Kriging
Log of distance to airport	X	X		Kriging
Log of distance to major road	X	X		Kriging
Log of distance to business centre	X	X		Kriging
Log of distance to railroad	X	X		Kriging
Log of distance to Tar River	X	X		Kriging
Log of distance to park	X	X		Kriging
Sold between Fran and Floyd (=1)	X	X		
Sold after Floyd (=1)	X	X		
Floodplain (=1)	X	X	X	X
Floodplain × sold btw Fran and Floyd	X	X		
Floodplain × sold after Floyd	X	X		

A variable ‘Schools’ has not been included in the analysis since the school quality in this particular areas is rather homogeneous. Furthermore, school rating does not have statistically significant effect on property prices

$$\ln P^i = \beta_0 + \sum_{k=1}^K \beta_k x_k^i + E^i \tag{2}$$

where $\ln P^i$ is the natural log of the sales price of property i , β_0 is the intercept, β_k is the coefficient for each property characteristic k , x_k^i is the value of characteristic k of property i , and E^i is the residual of the predicted property price. Our objective is to identify the model that provides the smallest prediction errors in the out-of-sample predictions for various sample sizes. The models to be compared include:

- A spatial hedonic model from Bin and Landry (2013) (M1)
- An adjusted version of M1 with different functional forms (M2)
- M2 with a reduced number of input variables (M3)
- M3 whereby spatial variability in property prices is predicted with kriging (M4)

The hedonic analysis is used to estimate the coefficients of the input variables β_k (Eq. 2). In M4 hedonic analysis is used to understand the influence of the core spatial variable of interest (flood risk) on property prices while the rest of the spatial variability in prices is captured by interpolating the residuals (E^i , Eq. 2) using kriging. A list of all the input variables can be found in Table 2. M1 is same model with the same specifications as used

in Bin and Landry (2013). However, the model in Bin and Landry (2013) has more dummy variables than M1 in this paper, as they distinguish the 100 and 500-year flood zones. The reason this separation is not made in M1 is that only 2% properties in the 500-year floodplain were sold, so that the 500-year floodplain properties are many times absent in this subset. Therefore, we merge the 100 and 500-year flood zone properties in one floodplain variable.

M2 was built with the same input variables as M1, while changing some of the functional forms in order to better describe the saturation behaviour of the variable's influence on price. The variables $bedrooms^2$, age^2 , $square\ footage^2$ and $acres^2$ were substituted by $\sqrt{bedrooms}$, \sqrt{age} , $\ln(square\ footage)$ and \sqrt{acres} respectively. These variables often enter the hedonic analysis function in the quadratic specification (Case et al. 2004). Yet, in our dataset price dependence on them does not necessary follow the parabolic form (Do and Grudnitski 1993; Goodman and Thibodeau 1995) (Fig. 5, Appendix 1). In M3 and M4 we consider a reduced regression that contains only the main characteristics of the properties—sq. footage, bedrooms, acres and age—that have a clear, straightforward and always statistically significant effect on price.²

In M4, the kriging procedure compliments the hedonic analysis to explain spatial correlation in property prices. The hedonic regression of M3 does not contain spatial characteristics of the property. All spatial relations and neighbourhood effects are captured by the residuals in predicted property prices, E^i (Eq. 2). Therefore, E^i in the out-of-sample prediction is added as a function of the in-sample residuals E^j of nearby properties (Eq. 3). This function can be written as

$$E^i = \sum_{j=1}^N W^{i,j} E^j,$$

$$\sum_{j=1}^N W^{i,j} = 1 \quad (3)$$

whereby $W^{i,j}$ is a spatial weight matrix which specifies how much the residual price of an in-sample property j affects the residual of an out-of-sample property i , and depends on the spatial distance between property i and property j . The spatial weight matrix is derived from a model of variance as a function of distance, called a semivariogram. The semivariogram is constructed by calculating the variance in E for all point pairs within a certain distance class. It is expected that this variance increases when the distance between properties increases, but that it levels off with increasing distance. This relation is fitted with an exponential function, which is used to specify the spatial weight matrix; nearest properties get the highest weight. The user can specify how many of the nearest properties are taken into account in the kriging interpolation, which is a trade-off between computation time and prediction accuracy (see Alemi et al. 1988; Delhomme 1978 for more details on the kriging procedure). In our case we take the 15 nearest properties.³

² We did a thorough analysis with various combinations of log and square root functional forms to identify the functional forms that best fitted the transaction data.

³ We did a sensitivity analysis and concluded that 10–20 nearby properties is a good number to use since it does not change the model's performance. More than 20 does not change the performance but enhances computation time, while less than 10 properties reduces the prediction performance. The results of our sensitivity analysis are available upon request.

2.3 Analysis

We split the entire data set (4779 observations) into in-sample and out-of-sample parts. The coefficients of the hedonic regressions are calculated using in-sample data, which is constructed as a subset of the sales data ranging from 0.1 to 20%. We systematically vary this share to understand how small the in-sample subset can be to be able to deliver an acceptable predictive power for each of the models. The remaining transactions form the out-of-sample dataset, which we use to compare the prices predicted by the four models against the actual sales price. The coefficients of the in-sample hedonic regressions are used to form the predicted prices of the out-of-sample properties in M1, M2 and M3. In M4 we sum the regression estimates and the kriged residuals.

Further, we apply Monte Carlo method by taking 50,000 random subsets ranging from 0.1% ($N = 48$) to 20% ($N = 956$) of the dataset, with constant density.⁴ Each model is calibrated with the same subsets, so that we can make pairwise comparisons. There is no limit to the number of subsets we can take, so we decide to take enough to cover the full range of model performances. The performance of a model may strongly depend on the subset it is calibrated on. Therefore we choose to look not only at average performance, but also at the 95% confidence interval of the performance range across sample sizes.

First, M1 and M2 are compared to assess how changing functional forms affect the prediction performance across sample sizes. Second, M2 and M3 are compared to assess the effect of reducing the number of input variables in the hedonic analysis, which should be more suitable for predicting prices based on small samples. Third, M3 and M4 are compared to see the effect of kriging for different sample sizes. And finally, the performances of all models are compared to see which model performs best across various sample sizes.

For comparing the models' prediction performances, we use the following metrics: Root Mean Squared prediction Error (RMSE) (Bin 2004; Case et al. 2004; Selim 2009), Mean Absolute prediction Error (MAE) (Bin 2004; Case et al. 2004; Selim 2009), Standard Deviation of prediction Error (SDE) (Case et al. 2004) and Adjusted R-squared (Laurice and Bhattacharya 2005). We look at the Adjusted R^2 of the regression described by

$$\ln(\text{actual value}) \sim \ln(\text{predicted value}) \quad (4)$$

which is a measure of the model's explanatory power. RMSE, MAE and SDE are measures for prediction accuracy and precision (see Appendix 2).

3 Results

3.1 Comparing Functional Specifications of the Full Hedonic Model (M1 and M2)

Comparing M1 and M2 (Fig. 1), the prediction performance of M2 is considerably better than M1, with the strongest effect at small sample sizes. Looking at MAE, we see that the median performance of both models does not differ at sample fractions above 0.10 ($N > 480$). However, when looking at the metrics RMSE and SDE the median prediction performance of M1 is still worse than M2 at samples of $N \approx 240$. RMSE and SDE are performance metrics that 'punish' the model for highly inaccurate predictions. We see that when price assessments for the out-of-sample set need to be made on a rather small in-sample dataset M1 produces highly volatile prediction outputs, with price predictions that sometimes deviate from actual

⁴ The analysis was done in R, version 3.2.0. Computation time was approximately 1 h.

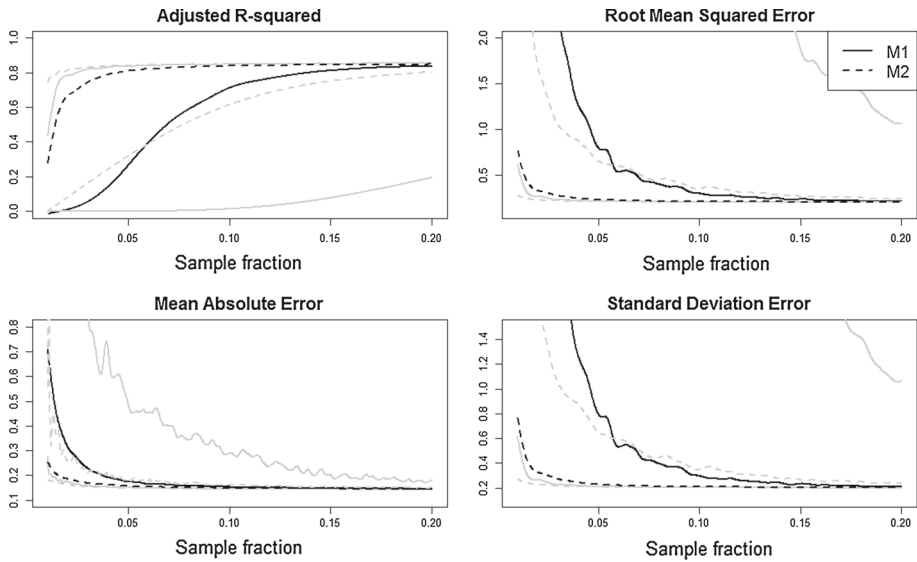


Fig. 1 Four out-of-sample prediction performance metrics as function of sample fraction, comparing M1 (continuous) and M2 (dashed). In black smoothing spline of the moving-window median. In grey the 95 % confidence interval boundary

sales by several orders of magnitude. The upper limits of the 95 % confidence intervals of M1’s performance metrics (Fig. 1) show that M1 often performs much worse than M2 as a result of these highly inaccurate price predictions.

The difference in performance between M1 and M2 can be explained by the functional forms of the input variables. The functional forms of some input variables of M1 are described by a squared relation with price, which is meant to represent the saturation behaviour of the characteristic’s effect on price. However, a squared or second degree polynomial function is not a saturating function. Rather, it has a peak (minimum value or maximum value depending on the sign of the coefficient) and can only approximate saturation behaviour on a local scale, but it can neither describe nor predict actual saturation behaviour overall. This results in a low model fit, which is measured by the Adjusted R-squared metric (Fig. 1). When the properties in the subset only contain a limited range of the characteristics compared to their entire range within the population, it can lead to large errors in the predicted price of properties with characteristics on the extreme ends of the range. Changing the squared functions to square root and log functions as in M2 could therefore considerably reduce the extreme prediction errors.

3.2 Reducing the Number of Explanatory Variables (M2 and M3)

Comparing M2 and M3 (Fig. 2) we see that M3 performs consistently better at sample fractions below 0.05 ($N < 240$). The prediction performance of M3 measured by all 4 metrics is relatively constant across sample sizes, whereas M2’s performance reduces drastically with decreasing sample sizes. This shows that a decrease in the number of variables in the hedonic analysis can enhance the model’s prediction performance when it is calibrated with few sales. Over-fitting is likely to happen in the case of small sample sizes and large numbers of explanatory variables. In these cases, the explanatory variables that have little explanatory

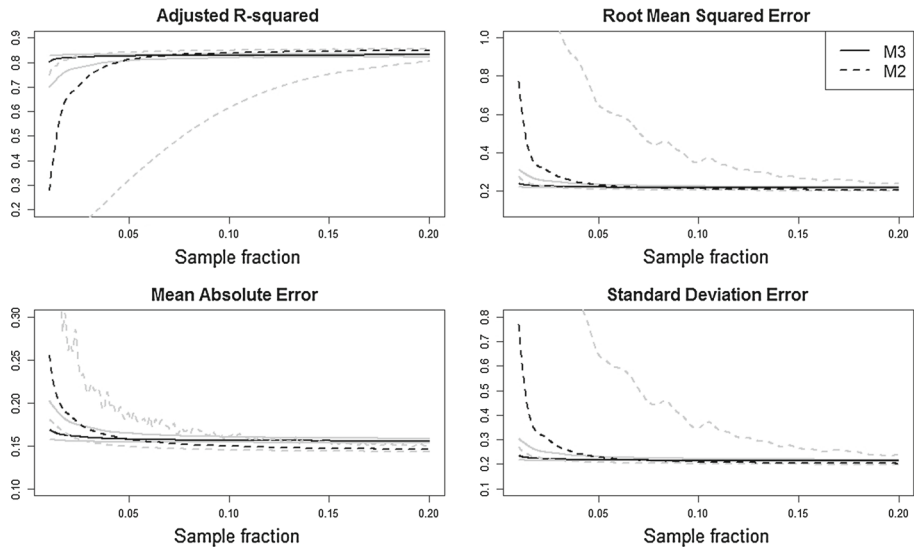


Fig. 2 Four out-of-sample prediction performance metrics as function of sample fraction, comparing M3 (continuous) and M2 (dashed). In black: smoothing spline of the moving-window median. In grey: the 95 % confidence interval boundary

value will not contribute to the prediction performance of the model, but rather start explaining noise within the sample. This results in a loss of generalisation in the model, and an increase in the stochastic behaviour of the predicted values. It is therefore crucial that the hedonic model only takes into account the factors that are of key importance and have a clear and sensible effect on the price, especially when it is based only on few observation data points.

At sample fractions higher than 0.05 (or $N > 240$) M2 generally has a lower mean absolute error than M3, suggesting that M2 outperforms M3 when they are calibrated with samples of $N > 240$. However, this conclusion does not hold when looking at the other performance metrics. The upper part of the 95 % confidence interval of RMSE and SDE shows that the predictions of M2 can still be quite volatile compared to M3. Thus even though M2 performs better than M3 on average with samples of $N > 240$, the precision of M2’s predictions is still lower than that of M3, where M2 has a higher probability of strongly inaccurate predictions.

3.3 Explaining Spatial Variability in Prices Through Kriging (M3 and M4)

From the first glance the previous discussion suggests that dropping various spatial factors when explaining property prices may be attractive. However, this is relevant only for small samples and serves as a disadvantage for larger samples. Indeed, properties in the same neighbourhood share similar spatial attributes such as proximity to parks, highways or other transport hubs, shopping centres, schools, and so on. It is a loss not to control for these given data availability. At the same time, the influence of all of them is not really vital for the research question at hand. Often one only needs to zoom in into a few specific attributes explaining price variations in the market and exclude the rest in an attempt to improve hedonic model performance. As we have seen above, including too many explanatory factors may jeopardize the latter for relatively small in-samples.

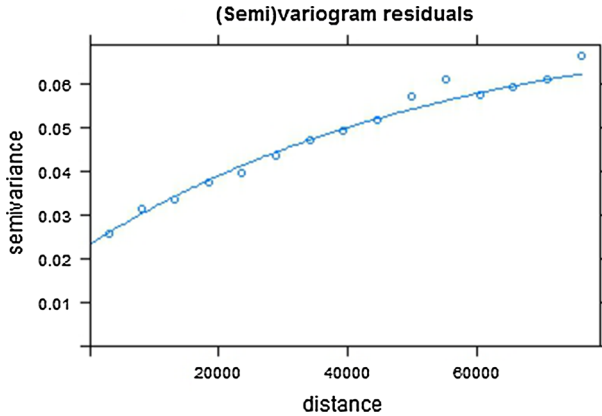


Fig. 3 Semivariogram of the in-sample residuals of M3. The *graph* shows that there is spatial correlation in the residuals, since the variance in residuals increases with distance between properties

Kriging offers an alternative way to account for the influence of spatial complexity in price assessments. Namely, it captures any systematic variation in prices through the analysis of residuals. In our dataset we find a clear spatial correlation in the semivariogram of the in-sample residuals of M3 (Fig. 3). The semivariance increases with distance between properties, which shows that property prices are spatially correlated. This indicates that the prediction performance of M3 can be improved with regression kriging (Basu and Thibodeau 1998). Regression kriging is done with the same variables as the hedonic model M3, after which the residuals are interpolated, so that only the remaining variation is addressed.

Comparing M3 (reduced hedonic model without kriging) and M4 (with kriging), we see that the model's performance consistently improves when kriging captures the spatial autocorrelation in residuals (Fig. 4). SDE, MEA and RMSE are consistently lower when kriging is added to M3 (Wilcoxon Signed-Rank Test, $P < 0.001$), and Adjusted R-squared is consistently higher with kriging (Wilcoxon Signed-Rank Test, $P < 0.001$). Most importantly, this result is consistent across all sample fractions.

The enhanced performance of M4 with kriging is more pronounced with increasing sample sizes (Fig. 4). At small sample sizes the density of the sample is so low that some of the predicted property prices are not spatially correlated with any of the property prices in the sample. The chance of having out-of-sample property prices that are spatially uncorrelated with the sample diminishes when the density of the sample increases. This is why the improvement of kriging becomes more pronounced with increasing sample sizes.

When comparing the four models, we find that M4 with kriging performs consistently better across various sample sizes. Thus, it delivers a more robust model to be used in hedonic analysis without a need for a researcher to worry about meeting a particular threshold of an in-sample size: it simply performs well for a large variety of sample sizes. Looking at the performance metrics, M3 with kriging is the best model in 96.2% of the cases with RMSE and SDE, 97.2% of the cases with MAE and 98.9% of the cases with Adjusted R-squared. It also implies that kriging explains the spatial variability in property prices better than the spatial variables that are included in M1 and M2 and releases a researcher from worrying about functional specifications of the spatial variables of secondary importance. At the same time, using kriging in combination with traditional hedonic analysis allows disentangling

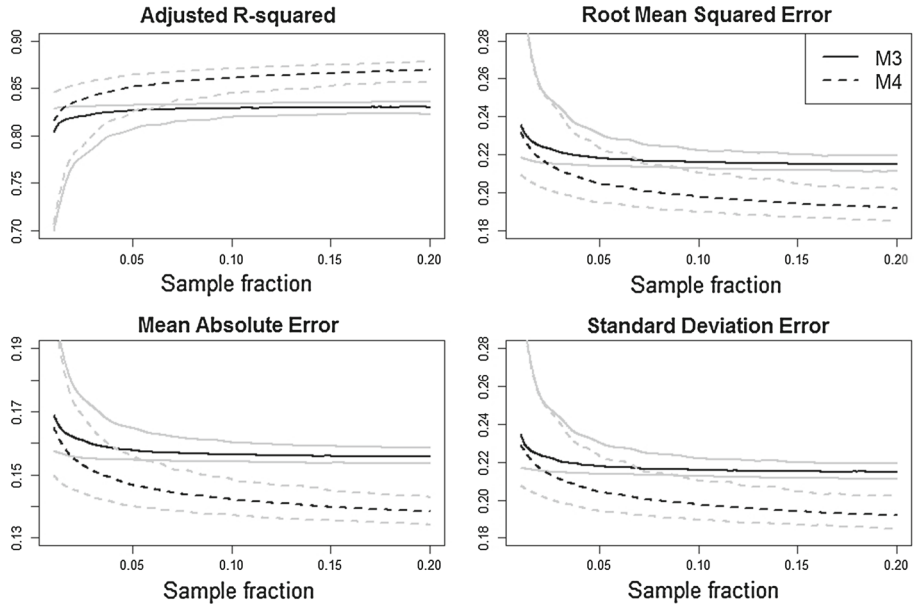


Fig. 4 Out-of-sample prediction performance of M3 (continuous) and M4 with kriging (dashed) as function of sample fraction. In black: smoothing spline of the moving-window median. In grey: 95 % confidence interval boundary

spatial attributes of a particular interests for economic analysis—e.g. a location within a flood zone in our case—to be studied with precision.

3.4 Implications for Policy Making: Example of the Flood Risk Discount

Outcomes of a hedonic analysis often serve as inputs for a larger CBA. When discussing flood risk management policy, either an adaptive estimation of the capital at stake (i.e. overall property price assessment) or a particular value of a flood risk discount (i.e. a value of the regression coefficient for a flood-zone dummy) plays a major role in the estimation of costs and benefits of a particular measure. Yet, what does the sensitivity to the in-sample size imply for a CBA? Let us examine the flood risk discount in particular (Table 3). It must be noted that the flood-zone dummy in M1 and M2 is also represented by two interaction terms (between hurricanes Fran and Floyd, and after hurricane Floyd) to see how the flood coefficient changes over time. Thus, the flood risk coefficient for M1 and M2 in Table 3 was controlled for the interaction terms. Table 3 presents the averages across the 50,000 Monte Carlo sampling runs.

The full regression model with the traditional functional specification M1 provides a rather unstable estimation of the flood risk coefficient since it varies greatly with the sample size. In fact for small in-sample sizes the flood dummy coefficient is positive but its standard deviation across random Monte Carlo sampling sets is huge undermining its statistical insignificance. The same can be said about M2 for small in-sample sizes, although the standard deviation is already much lower than M1. M3 and M4 have the lowest standard deviation for small in-sample fractions (<0.05). With larger sample fractions (>0.10) it does not matter anymore which model is used to predict the flood coefficient.

Table 3 Stability of a regression coefficient of the floodplain variable over various in-sample sizes

Sample fraction	M1		M2		M3 & M4	
	Mean	SD	Mean	SD	Mean	SD
<0.05	0.150	17.034	-0.055	0.447	-0.050	0.109
<0.10	-0.063	0.103	-0.061	0.056	-0.052	0.055
<0.15	-0.061	0.040	-0.061	0.041	-0.053	0.041
>0.15	-0.060	0.033	-0.060	0.033	-0.053	0.033
Total population	-0.056	-	-0.056	-	-0.054	-

Results in Table 3 suggest that the sales price differential between inside and outside the floodplains ranges from 5.0 to 6.3 % with an exception of M1 with less than 0.05 sample fraction. Several previous studies have documented the price reduction from location in a floodplain (MacDonald et al. 1987; Bin and Polasky 2004; Hallstrom and Smith 2005; Bin et al. 2008; Daniel et al. 2009). A common finding in these studies is that location within a floodplain lowers property value anywhere from 4 to 12 %. As shown in Table 3, our approach to limit the number of variables that enter the hedonic regression can be quite useful in determining the risk premiums associated with flooding especially with small samples. Our results may help insurance practitioners and policy makers make informed decisions on the flood risk management especially when the available data set is very limited.

4 Discussion

Across all sample sizes we see that M4 with kriging performs best in the out-of-sample predictions regardless of the in-sample size. This model differs from M1 and M2 in the way the spatial variables enter the price estimation. In M3 the spatial variables were omitted completely, whereas kriging was used in M4 to predict spatial variability and spatial autocorrelation in property prices by analysing the residuals. Kriging-based M4 is thus more powerful in predicting the spatial patterns in property prices. This may not be surprising, as kriging is used for spatial interpolation by explicitly accounting for spatial autocorrelation, which is often present in the property market (Basu and Thibodeau 1998; Bourassa et al. 2007; Dubin 1992; Militino et al. 2004). Case et al. (2004) have already shown that kriging can, for this reason, enhance the out-of-sample prediction performance of spatial hedonic models. Yet, their models were calibrated with very large samples ($N \approx 50,000$) and the question remained whether these conclusions hold for small samples. We have specified a model that is consistent in assessing housing prices and predicting future sales prices, even when calibrated with a limited number of recent sales. We can zoom into the mechanisms of why this model performs best by comparing M1, M2, M3 and M4 with kriging pairwise across different sample sizes.

Comparing M1 and M2 reveals that the squared terms in the hedonic model can cause large errors in out-of-sample predictions of the property prices, with estimated prices that sometimes deviate even by several orders of magnitude from the actual price. This problem especially occurred at small sample sizes for which M1 was calibrated. The poor out-of-sample prediction performances were expressed by a high variability in prediction errors and a low model fit. The latter indicates that the chosen functional forms of the variables may not represent their actual effect on price. Squared functional forms, and sometimes even cubed

functional forms, are currently widely used in hedonic literature (Bin and Landry 2013; Case et al. 2004). These models are based on datasets that are usually large enough to sanitise the effects. Yet, we found that even with samples of $N = 900$ the predictions can be inaccurate in the out-of-sample predictions as a result of a low model fit.

A comparison between M2 and M3 reveals that over-fitting is the main cause of the poor out-of-sample prediction performance of M2 at very small sample sizes. A similar mechanism called over-training is also causing poor prediction performances of artificial neural network models (Nguyen and Cippis 2001). Reducing the number of input variables in M3 results in a consistent prediction performance across sample sizes. Moreover, it leads to a decrease of volatility in the model's predictions. However, we also observe that M2 scores better on the metric mean absolute error when a sample size increases. This implies that some of the model's parameters can only be estimated when the number of sales is high. When trying to calibrate a model based on only a few recent sales, it is better to focus only on a few explanatory variables.

M4 with kriging outperforms M3 across the range of sample sizes, but performs only slightly better than M3 with sample sizes of $N \approx 50$. In fact, the improvement of kriging with respect to the prediction performance of M4 increases with sample sizes. The strong prediction performance of M3 and M4 at small sample sizes is mainly due to the reduction of variables in the hedonic model, whereas the role of kriging in improving the prediction performance of M4 becomes more important with increasing sample sizes. The latter is caused by an increase in density of the sample, so that the 15 nearest properties that are selected for spatial interpolation are on average closer to the predicted location, i.e.: their actual transaction prices are more correlated with the predicted price. To summarize: we observe that changes in functional forms of the input variables, a reduction of input variables and kriging improve the prediction performance at different parts of the range of sample sizes. Together, these specifications complement each other to form a model that is consistently better in prediction performance across sample sizes.

Despite that M4 consistently produces the best price predictions across a range of sample sizes, we do not suggest that it can replace the hedonic analysis when it is used for other purposes, namely to assess marginal implicit prices of specific housing attributes (Janssen et al. 2001). In this case the spatial attributes of interest should be kept in the hedonic function part of the analysis while the impact of other spatial neighbourhood attributes on price may be captured by kriging. The limitation of kriging is that all involved spatial attributes go into the black box of spatial interpolation. Thus, one cannot trace back which spatial factors exactly affect property prices and to what extent.

For the purpose of predicting property values, for example in real estate appraisal (Pagourtzi 2003), it can be useful to work with models that are consistent in their performance even when calibrated with few current sales. Our model, which performs consistently well across a range of sample sizes, is particularly useful for this application. For policy makers that deal with management of natural hazards it is important that a good assessment of the capital at risk is made. Housing markets are affected by macro-economic changes as well as changes in consumer preferences, incomes and WTP for various property attributes. Housing markets in hazard areas experience structural changes in price trends after disastrous events, and these changes are expected to accelerate with climate change. Thus, when conducting a CBA for flood risk management policies it is essential that the assessment of capital at risk or a flood risk premium is based on the most recent sales to better reflect current market conditions. For example when dealing with flood risk, it is important to account for changing risk perceptions, which is driven by flood events and changing flood probabilities, influencing how risk is capitalized into property prices (Pryce et al. 2011; Atreya et al. 2012; Bin

and Landry 2013). Our algorithm allows for a rapid updating of property price assessments and predictions. Therefore, it can quickly capture a market response to potential changes in location preferences, market conditions and flood risk perceptions. This approach can be used to monitor price changes in risk-prone areas, accounting for changes in flood risk and at the same time controlling for autonomous market responses to flood risk.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix 1

See Fig. 5.

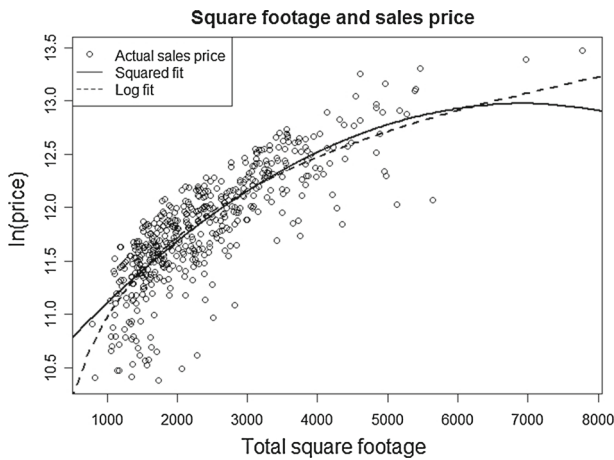


Fig. 5 Relationship between house square footage and sales price, including the fitted regressions with squared and log functional forms

Appendix 2

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_{pred} - Y_{obs})^2}{n}} \tag{5}$$

$$MAE = \frac{\sum_{i=1}^n |Y_{pred} - Y_{obs}|}{n} \tag{6}$$

$$SDE = \sqrt{\frac{\sum_{i=1}^n (Y_{pred} - Y_{obs})^2}{n - 1}} \tag{7}$$

References

- Alemi MH, Azari AS, Nielsen DR (1988) Kriging and univariate modeling of a spatially correlated data. *Soil Technol* 1:133–147
- Anselin L (2001) Spatial econometrics. In: Baltagi BH (ed) *A companion to theoretical econometrics*, chap 14. Wiley, Sons, pp 310–330
- Atreya A (2013) Flood risk, homeowners' flood risk perception and insurance. Dissertation, University of Georgia
- Atreya A, Ferreira S, Kriesel W (2012) Forgetting the flood: changes in flood risk perceptions over time. In: *Flood risk and homeowners' flood risk perceptions: evidence from property prices in Georgia*, chap 1. pp 5–36
- Basu S, Thibodeau TG (1998) Analysis of spatial autocorrelation in house prices. *J Real Estate Finance Econ* 17(1):61–85
- Bin O (2004) A prediction comparison of housing sales prices by parametric versus semi-parametric regressions. *J Hous Econ* 13:68–84
- Bin O, Kruse JB, Landry CE (2008) Flood hazards, insurance rates, and amenities: Evidence from the coastal housing market. *J Risk Insur* 75(1):63–82
- Bin O, Landry CE (2013) Changes in implicit flood risk premiums: empirical evidence from the housing market. *J Environ Econ Manag* 65:361–376
- Bin O, Polasky S (2004) Effects of flood hazards on property values: evidence before and after hurricane Floyd. *Land Econ* 80(4):490–500
- Bostic R, Gabriel S, Painter G (2005) Housing wealth, financial wealth, and consumption: new evidence from micro data. *Reg Sci Urban Econ* 39(1):79–89
- Bourassa SC, Cantoni E, Hoesli M (2007) Spatial dependence, housing submarkets and house price prediction. *J Real Estate Finance Econ* 35:143–160
- Case B, Clapp J, Dubin R, Rodriguez M (2004) Modelling spatial and temporal house price patterns: a comparison of four models. *J Real Estate Finance Econ* 29(2):167–191
- Daniel VE, Florax RJGM, Rietveld P (2009) Flooding risk and housing values: an economic assessment of environmental hazard. *Ecol Econ* 69(2):355–365
- Delhomme JP (1978) Kriging in the hydrosocieties. *Adv Water Resour* 1(5):251–266
- Do AQ, Grudnitski G (1993) A neural network analysis of the effect of age on housing values. *J Real Estate Res* 8(2):253–264
- Dubin RA (1992) Spatial autocorrelation and neighborhood quality. *Reg Sci Urban Econ* 22:433–452
- Dubin RA (1999) Predicting house prices using multiple listings data. *J Real Estate Finance Econ* 17(1):35–59
- Dutta D, Herath S, Musiak K (2003) A mathematical model for flood loss estimation. *J Hydrol* 277:24–49
- Farber S (1987) The value of coastal wetlands for protection of property against hurricane wind damage. *J Environ Econ Manag* 14:143–151
- Filatova T (2014) Empirical agent-based land market: integrating adaptive economic behavior in urban land-use models. *Comput Environ Urban Syst* 54:397–413
- Gamper CD, Thöni M, Weck-Hannemann H (2006) A conceptual approach to the use of cost benefit and multi criteria analysis in natural hazard management. *Nat Hazards Earth Syst Sci* 6:293–302
- Goodman AC, Thibodeau TG (1995) Age-related heteroskedasticity in hedonic house price equations. *J Hous Res* 6(1):25–42
- Hall JW et al (2003) A methodology for national-scale flood risk assessment. *Proc ICE Water Marit Eng* 156(3):235–247
- Hall JW, Sayers PB, Dawson RJ (2005) National-scale assessment of current and future flood risk in England and Wales. *Nat Hazards* 36:147–164
- Hallegatte S (2006) A cost-benefit analysis of the New Orleans Flood Protection System. In: AEI-Brookings joint center. Regulatory analysis, pp 06–02. <https://halshs.archives-ouvertes.fr/hal-00164628/>
- Hallstrom DG, Smith VK (2005) Market responses to hurricanes. *J Environ Econ Manag* 50:541–561
- Hernandez-Stefanoni JL, Ponce-Hernandez R (2006) Mapping the spatial variability of plant diversity in a tropical forest: comparison of spatial interpolation methods. *Environ Monit Assess* 117:307–334
- Hirabayashi Y et al (2013) Global flood risk under climate change. *Nat Clim Change* 3:816–821
- Janssen C, Söderberg B, Zhou J (2001) Robust estimation of hedonic models of price and income for investment property. *J Prop Invest Finance* 19(4):342–360
- Kuntz M, Helbich M (2014) Geostatistical mapping of real estate prices: an empirical comparison of kriging and cokriging. *Int J Geogr Inf Sci* 28(9):1904–1921
- Laurice J, Bhattacharya R (2005) Prediction performance of a hedonic pricing model for housing. *Apprais J* 73(2):198–209

- MacDonald DN, Murdoch JC, White HL (1987) Uncertain hazards, insurance, and consumer choice: evidence from housing markets. *Land Econ* 63:361–371
- Merz B et al (2010) Assessment of economic flood damage. *Nat Hazards Earth Syst Sci* 10:1697–1724
- Militino AF, Ugarte MD, García-Reinaldos L (2004) Alternative models for describing spatial dependence among dwelling selling prices. *J Real Estate Finance Econ* 29(2):193–209
- Nguyen N, Cippa A (2001) Predicting housing value: a comparison of multiple regression analysis and artificial. *Neural Netw* 22(3):313–336
- Oliveri E, Santoro M (2000) Estimation of urban structural flood damages: the case study of Palermo. *Urban Water* 2:223–234
- Pagourtzi E et al (2003) Real estate appraisal: a review of valuation methods. *J Prop Invest Finance* 21(4):383–401
- Penning-Rowsell E et al (2005) *The benefits of flood and coastal risk management: a handbook of assessment techniques*. Middlesex University Press, London
- Pryce G, Chen Y, Galster G (2011) The impact of floods on house prices: an imperfect information approach with Myopia and Amnesia. *Hous Stud* 26(2):259–279
- Selim H (2009) Determinants of house prices in Turkey: hedonic regression versus artificial neural network. *Expert Syst Appl* 36(2):2843–2852
- Voltz M, Webster R (1990) A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. *J Soil Sci* 41(3):473–490
- Ward PJ et al (2014) Strong influence of El Niño southern oscillation on flood risk around the world. *Proc Nat Acad Sci* 111(44):15659–15664
- Webster R, Burgess TM (1983) Spatial variation in soil and the role of kriging. *Agric Water Manag* 6:111–122