Spatial and hedonic analysis of house price dynamics in Warsaw

Widłak, Marta and Waszczuk, Joanna and Olszewski, Krzysztof

Narodowy Bank Polski

August 2014

Spatial and hedonic analysis of house price dynamics in Warsaw

Widłak Marta[1], Waszczuk Joanna[1], Krzysztof Olszewski[1]

August 2014

Summary

The aim of our article is to analyze the dynamics of housing prices in the secondary housing market in Warsaw from Q1 2006 to Q3 2013, taking into account the spatial relationship between prices. In the first part of this research we compare the geographically weighted regression with a linear regression estimated using OLS with spatial variables. In the second part, we combine the geographically weighted regression with the penalized spline regression to extract the effect of time on prices. With this method we obtain a nonlinear and more precise measure of time effects and improved goodness-of-fit statistics. We obtain a hedonic index, that is more robust against short-term changes in house prices than the usual, linear hedonic index. This is a novel approach, which has not been applied before in the case of the Polish housing market. The index allows us to show how interest rates or the housing policy influenced house prices.

JEL Classification Codes: D12, R21, R31.

Key words: hedonic price indices, housing prices, spatial influence on prices, geographically weighted regression, penalized splines

---

[1] Narodowy Bank Polski, Financial Stability Department (Świętokrzyska 11/21; 00-919 Warszawa, Poland).

Corresponding Author: Joanna.Waszczuk@nbp.pl

We would like to thank participants of the III Spatial Econometrics and Regional Economic Analysis Conference at the University of Łódź and the Warsaw International Economic Meeting 2014 for their hints that helped us to improve this paper.

1. Introduction

The housing market plays an enormous role for every household and is an important sector that contributes to the growth of the whole economy. The purchasing of housing affects households in two ways, as they satisfy their housing needs and obtain housing stock. The world's recent experience and the crisis in the US showed that price fluctuations and the cyclical character of the housing market are risk factors that can threaten the stability of the financial and banking sectors. The accumulation of risks can lead to uncertainty in the market that can worsen the economic situation because of links between the housing, banking and finance sectors. Price changes can have an impact on the whole country and this is the reason why central banks and regulators monitor house price dynamics. When measuring house price fluctuations, it is worth remembering about the heterogeneity of housing units, since the price movements can be caused by differences in housing attributes in the analyzed sample. To capture this we applied hedonic models.

Hedonic models explain the house price using its characteristics, such as the area, standard, and localization. The basis of this method is the "new theory of consumer demand" presented by Lancaster in 1966, and then developed by Rosen in 1974. Rosen claimed that the estimation of the value of particular attributes indirectly carries information about the outcome of supply and demand changes.[2] These are influenced by the expectations of consumers and sellers.[3]

In this article we stress the influence of spatial dependence and housing attributes on house prices fluctuation. The beginnings of spatial analyses come from the 1970s, when Tobler formulated the first law of geography, stating that "Everything is related to everything else, but near things are more related than distant things." Well-known scientists in the field of spatial econometrics are, among others, Moran, Geary, Paelinck and Anselin.

There are only a few publications concerning spatial modelling in Poland, and in the case of housing they are confined to just a few articles. Dale-Johnson et al. (2001) used the spatial regression to discover how the price of vacant commercial lots in Kraków depends on

[2] Works in Poland in the field of hedonic models are among others: Tomczyk, Widłak (2010); Widłak, Nehrebecka (2011); Widłak (2013).

[3] The analysis of consumer behavior in housing sector is presented in Łaszek 2013.

their characteristics and variables defining its localization. Dale-Johnson et al. (2005) continued to analyse lots in Kraków, emphasizing the asymmetry of the geographical distribution of prices around many local maximum points of housing values. Moreover, they include some variables that described the distance from the building to the point with the highest price and they conclude that the more centres they take into account, the better results they obtain. Kulczycki et al. (2007) use the geographically weighted regression to explain the value of housing units in one of Kraków's districts - Krowodrza from 2004 to 2005. Cellmer (2010) analysed the price dynamics in Olsztyn using the same method. Branna et al. (2012), obtained, among others, geographically weighted regression and a hedonic model estimated by OLS containing some spatial variables for Kraków. Waszczuk (2013) analysed the price of housing in Warsaw using the spatial lag model (SLM), the spatial error model (SEM) and the geographically weighted regression. It turned out that the geographically weighted regression gives the most promising results.

It should be noted that the development of spatial models in Poland in the field of housing markets is slow because of the lack of data that describes the housing unit localization or its attributes in detail and because access to information that some public or business entities have in their possession is restricted. This causes market intransparency and results in a poor quality of the outcome of many analyses that are based on incomplete information, reflecting only a small part of market variability.

In this article we use the iterative penalized spline regression, like Brunnauer et al. (2011) and Menedez et al. (2013), to estimate the hedonic index. This method is better than the standard hedonic index, as it describes price trends in a better way than a linear model, especially when there are a short-term fluctuations in prices.

Our article consists of 4 chapters. In chapter 2 we discuss our data set and we present the results of a hedonic model estimated by OLS and the geographically weighted regression. In chapter 3 we estimate a price index using the iterative penalized spline regression and analyse the economic and political factors that influence price movements. In chapter 4 we conclude the paper.

2. House price models

In this chapter we present two models in which the price of a flat is explained by its attributes and its localization. We use information about transactions from the secondary housing market that are gathered in the BaRN data base.[4] Our data set consists of physical attributes of dwellings, such as the number of rooms, standard and many more mentioned in appendix 1. We geocode (determine the latitude and longitude coordinates)[5] the observations and computed "geovariables" (spatial variables), that can potentially influence the price of housing units. To minimize the errors in the estimations, we decided to exclude dwellings located in streets that are longer than 3 kilometres. The next step was to clear the data set and reject transactions, if the studentized residuals are equal to or more than 2.5. The final data set includes 57 independent variables and 1 dependent variable (the log price) and consists of 4037 transactions concluded from Q1 2006 to Q3 2013.

2.1 Hedonic model estimated by OLS

The first model (**M1**) is estimated by OLS, including some spatial variables: district-dummies, the logarithms of the distance from the metro station, the distance from the city centre, and dummies determining if there are green areas nearby. Tests indicate that the model is specified correctly (we carried out the RESET test) and the collinearity is moderate (VIF<10). We use the heteroskedasticity and autocorrelation consistent estimator (HAC). On the basis of our estimation results, presented in table 1, we confirm that most variables are statistically significant and their estimated direction of influence on house prices is as we conjectured.

Table 1. The result of the ordinary least squares model - M1.

Variable	Coef.	Std. Error	t value	
Intercept	9.6889	0.0648	149.6400	***
l_stacja_m	-0.0217	0.0023	-9.5600	***
zielen_1	-0.0816	0.0133	-6.1500	***
l_centrum_dist	-0.0804	0.0075	-10.7400	***

[4] Database created by Narodowy Bank Polski.

[5] In most cases we geocode on the basis of the middle of the street because we do not possess information about building numbers.

powierzchnia	-0.0048	0.0004	-12.2000	***
sq_pow	0.0000	0.0000	9.5300	***
ilepokoi_1_2	0.0142	0.0060	2.3600	*
stand_1	0.0629	0.0049	0.0049	***
stand_3_4	-0.0467	-0.0467	-0.0467	***
spoldzielcze	-0.0166	0.0045	-3.6700	***
rok_1_2	-0.0251	0.0130	-1.9300	*
rok_3	-0.1100	-0.1100	-13.8300	***
rok_4	-0.1314	0.0075	-17.4200	***
rok_5	-0.1435	0.0077	-18.6700	***
rok_7	0.0648	0.0078	8.2800	***
rok_8	0.1031	0.0096	10.7800	***
rok_nowe	-0.0213	0.0156	-1.3700	
gr1	-0.1311	0.0154	-8.5100	***
gr2	-0.2730	0.0173	-15.7700	***
gr3	-0.1581	0.0146	-10.8200	***
gr4	-0.3244	0.0239	-13.5500	***
gr5	-0.1790	0.0189	-9.4800	***
mokotow	-0.0435	0.0135	-3.2300	***
ochota	-0.0630	0.0141	-4.4700	***
wola	-0.1251	0.0105	-11.8900	***
ursynow	-0.0409	0.0175	-2.3400	**
zoliborz	-0.0515	0.0165	-3.1100	***
wilanow	-0.0767	0.0226	-3.3900	***
t_20062	0.0982	0.0214	4.5800	***
t_20063	0.2320	0.0204	11.3500	***
t_20064	0.3995	0.0220	18.1800	***
t_20071	0.4822	0.0213	22.6300	***
t_20072	0.5593	0.0213	26.3100	***
t_20073	0.6205	0.0214	29.0400	***
t_20074	0.6526	0.0237	27.5400	***
t_20081	0.5828	0.0219	26.5600	***
t_20082	0.5770	0.0210	27.4900	***
t_20083	0.5819	0.0229	25.4300	***
t_20084	0.5790	0.0225	25.7500	***
t_20091	0.5239	0.0249	21.0600	***
t_20092	0.5377	0.0236	22.7900	***
t_20093	0.5112	0.0231	22.1200	***
t_20094	0.5223	0.0224	23.3100	***
t_20101	0.5383	0.0217	24.7800	***
t_20102	0.5660	0.0212	26.7500	***
t_20103	0.5528	0.0215	25.7300	***
t_20104	0.5178	0.0205	25.2200	***
t_20111	0.5211	0.0202	25.8500	***
t_20112	0.5316	0.0187	28.4200	***
t_20113	0.5369	0.0187	28.7400	***
t_20114	0.5262	0.0192	27.4100	***
t_20121	0.5166	0.0192	26.9100	***
t_20122	0.4900	0.0190	25.7300	***
t_20123	0.4767	0.0202	23.6400	***
t_20124	0.4459	0.0191	23.3900	***
t_20131	0.4265	0.0216	19.7600	***
t_20132	0.4042	0.4042	19.5200	***
t_20133	0.4248	0.0187	22.7300	***

The coefficient of determination R^2 amounts to about 72%. The estimated values for particular districts are negative, as the reference variable for the dummy variables is the city centre Śródmieście (the most expensive district). For example, in the case of Wawer and Rembertów, districts that are located far away from city centre, dwelling prices are about 32% lower in comparison with those in Śródmieście. An increase in the distance from the city centre and metro station causes a fall in the value of flats, which is in line with the monocentric city model. Also if the dwelling is situated near green areas, the price will be higher. We observe that housing units that are newer cost more than comparable flats that were built earlier. An analysis of housing attributes shows that there is a non-linear, u-shaped relation between the size and the price of a flat. Small dwellings are more expensive than medium-sized ones and the value of a sq. m. starts to increase again in the case of bigger housing units. Firstly, small and very large dwellings are scarce goods. Moreover, small dwellings have higher prices per sq. m. but their overall value is lower, so they are more accessible for even poorer households, thus the demand for such flats is high.

In order to make the effect of distance to the city centre or the next metro station more linear, we apply the natural logarithm to these variables. Moreover, we analyze if the influence of other values on the price is non-linear. We use the penalized spline regression, which shows that the estimated function is not significantly different from a linear function in all cases. It is very important to add that in the geographically weighted regression we can only use linear elements, such as the size and squared size, but cannot use the penalized spline regression.

2.2 Spatial hedonic model – geographically weighted regression (GWR)

When we use the global econometric model for a city or a region, the assumption of spatial stationarity can be a problem. If the process that we examine is spatially heterogeneous, we usually need to use the geographically weighted regression. On the basis of the results of the geographically weighted regression we can analyze the spatial variability of the coefficients. Every observation is weighted by the matrix W_i and using the OLS method, we estimate the local values of the parameters. According to Suchecki (2010), the spatial weighting function has a very big influence on the results of the analysis.

The geographically weighted regression is described by the following equation:

$y_i = \beta_{i0}(u_i, v_i) + \sum_k \beta_{ik}(u_i, v_i) x_{ik} + \varepsilon_i,$

where:

y_i is the dependent variable,

(u_i, v_i) are the latitude and longitude coordinates,

$\beta_{ik}(u_i, v_i)$ is the realization of $\beta_k(u, v)$ in 'i' space point,

x_{ik} are independent variables,

ε_i are error terms.

The GWR parameters are estimated using weight matrixes (dependent on the localization) for every observation. We use the *adaptive bisquare* weighting function (kernel function) and our motivation is that the distribution of observations is unbalanced on the analyzed area. By using that particular weighting function, we avoid the situation that some dwellings on the outskirts of the city do not have any neighbours.

Table 2. The result of the geographically weighted regression - M2.

variable	Min.	1st Q	Median	3rd Q	Max.	OLS
Intercept,	9.4710	9.6240	9.6590	9.7180	9.8250	9.6889
l_stacja_m	-0.0309	-0.0258	-0.0195	-0.0170	-0.0152	-0.0217
zielen_1	-0.0908	-0.0841	-0.0720	-0.0581	-0.0524	-0.0816
l_centrum_	-0.0984	-0.0859	-0.0820	-0.0781	-0.0576	-0.0804
powierzchn	-0.0052	-0.0046	-0.0044	-0.0043	-0.0040	-0.0048
sq_pow	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ilepokoi_1	0.0035	0.0131	0.0148	0.0157	0.0197	0.0142
stand_1	0.0578	0.0641	0.0683	0.0729	0.0807	0.0629
stand_3_4	-0.0568	-0.0529	-0.0473	-0.0444	-0.0412	-0.0467
spoldzielc	-0.0268	-0.0203	-0.0162	-0.0131	-0.0067	-0.0166
rok_1_2	-0.0326	-0.0225	-0.0173	-0.0103	-0.0008	-0.0251
rok_3	-0.1187	-0.1060	-0.1015	-0.0933	-0.0804	-0.1100
rok_4	-0.1493	-0.1396	-0.1215	-0.1117	-0.0993	-0.1314
rok_5	-0.1591	-0.1501	-0.1289	-0.1161	-0.1120	-0.1435
rok_7	0.0386	0.0680	0.0820	0.0846	0.0934	0.0648
rok_8	0.0775	0.1042	0.1183	0.1335	0.1523	0.1031
rok_nowe	-0.0348	-0.0117	-0.0071	-0.0022	0.0002	-0.0213
gr1	-0.1345	-0.1224	-0.1156	-0.1013	-0.0639	-0.1311
gr2	-0.3138	-0.2484	-0.2279	-0.2196	-0.2035	-0.2730
gr3	-0.2175	-0.1540	-0.1481	-0.1428	-0.1312	-0.1581
gr4	-0.4516	-0.4312	-0.4134	-0.3351	-0.3068	-0.3243
gr5	-0.2371	-0.1483	-0.1291	-0.0976	-0.0454	-0.1790
mokotow	-0.0936	-0.0439	-0.0360	-0.0268	-0.0148	-0.0435
ochota	-0.1020	-0.0686	-0.0607	-0.0489	-0.0392	-0.0630
wola	-0.1568	-0.1342	-0.1263	-0.1117	-0.1042	-0.1251
ursynow	-0.0830	-0.0739	-0.0515	-0.0336	-0.0128	-0.0409
zoliborz	-0.0903	-0.0479	-0.0458	-0.0430	-0.0252	-0.0515
wilanow	-0.1311	-0.0833	-0.0786	-0.0459	-0.0144	-0.0767
t_20062	0.0818	0.0924	0.1016	0.1110	0.1161	0.0982

t_20063	0.2030	0.2269	0.2440	0.2486	0.2607	0.2320
t_20064	0.3704	0.3982	0.4171	0.4237	0.4312	0.3995
t_20071	0.4688	0.4825	0.4926	0.4986	0.5078	0.4822
t_20072	0.5468	0.5585	0.5678	0.5735	0.5802	0.5593
t_20073	0.6064	0.6206	0.6253	0.6316	0.6403	0.6205
t_20074	0.6046	0.6339	0.6656	0.6711	0.7045	0.6526
t_20081	0.5612	0.5689	0.5726	0.5864	0.6024	0.5828
t_20082	0.5549	0.5728	0.5814	0.5898	0.6090	0.5770
t_20083	0.5518	0.5652	0.5776	0.5841	0.6037	0.5819
t_20084	0.5669	0.5743	0.5769	0.5809	0.5952	0.5790
t_20091	0.5007	0.5208	0.5262	0.5309	0.5416	0.5239
t_20092	0.4976	0.5052	0.5339	0.5530	0.5624	0.5377
t_20093	0.4759	0.4909	0.5110	0.5156	0.5595	0.5112
t_20094	0.4907	0.5009	0.5230	0.5329	0.5770	0.5223
t_20101	0.5046	0.5425	0.5542	0.5565	0.5685	0.5383
t_20102	0.5435	0.5539	0.5673	0.5731	0.5915	0.5660
t_20103	0.5209	0.5350	0.5599	0.5682	0.5835	0.5528
t_20104	0.5011	0.5154	0.5175	0.5213	0.5304	0.5178
t_20111	0.5018	0.5183	0.5330	0.5402	0.5457	0.5211
t_20112	0.5129	0.5184	0.5353	0.5372	0.5582	0.5316
t_20113	0.5258	0.5367	0.5379	0.5405	0.5474	0.5369
t_20114	0.5090	0.5173	0.5262	0.5333	0.5443	0.5262
t_20121	0.4991	0.5068	0.5253	0.5292	0.5364	0.5166
t_20122	0.4715	0.4841	0.4887	0.4953	0.5187	0.4900
t_20123	0.4521	0.4680	0.4846	0.4896	0.5016	0.4767
t_20124	0.4234	0.4365	0.4388	0.4442	0.4709	0.4459
t_20131	0.4111	0.4174	0.4346	0.4359	0.4490	0.4265
t_20132	0.3754	0.3947	0.4151	0.4166	0.4293	0.4042
t_20133	0.4061	0.4182	0.4314	0.4343	0.4515	0.4248

We present the range of spatial variability of the estimated parameters in table 2. The results are satisfactory, as the global R^2 is higher and amounts to about 74% and the Akaike information criterion is lower than in model 1. In order to compare the results of the geographically weighted regression and the OLS model, we present the spatial distribution of residuals in the second part of the appendix (Figures 3, 4). Both models have relatively low residuals, but we can see that in the case of the GWR the residuals are slightly lower. The residuals have the highest values in the city centre, which means that the price over there is overestimated.

3. A spline-smoothed hedonic price indes

Following the work of Brunauer et al (2012) and Menedez et al. (2013) we apply penalized spline smoothing in order to obtain a hedonic price index. For a literature overview on the application of non-parametric estimation methods of hedonic models we refer to Brunauer (2010). We give a brief introduction to penalized spline smoothing and present the price index for Warsaw.

3.1 Brief introduction to spline smoothing

The penalized spline regression is a nonparametric estimation method using smooth functions which are not specified parametrically. The main idea is that the effect of an explanatory variable on a dependent one is not any more modelled by a linear function but by a flexible smoothing function. Therefore the effect can vary with the observations and this is represented by the spline. We only need to assume that the spline which models the effect is a continuous and differentiable function. The exact shape of this function is to be estimated from the data by use of spline smoothing.

The mathematical idea of splines was created by Reinsch (1967), but the milestone in the development of semi- and non-parametric estimation was set by Hastie and Tibshirani (1990), who introduced the Generalized Additive Model (GAM). Wood (2001, 2014) implemented the mgcv package in R, which allows for a free and quite simple usage of this estimation method. An overview of its application can be found in Venables and Ripley (2003), Ruppert et al. (2003), Greiner (2009), Greiner and Kauermann (2007).

We give a brief introduction following closely Greiner (2009). Assuming that we have n data points for a dependent variable y_t which is explained by the independent variable x_t, this leads to n observations of the following kind (y_t, x_t). The regression model we want to estimate is:

$$y_t = f(x_t) + e_t, \qquad e_t \sim iid(0, \sigma^2)$$

We do not specify $f(\cdot)$, but require it to be continuous and sufficiently differentiable. The idea is now to find a function which estimates our data in the best way but is not too wiggly and complex. Therefore the following minimization problem has to be solved (Hastie and Tibshirani (1990).

$$\min_{f(\cdot)}\left\{\sum_{t=1}^{n}(y_t - f(x_t))^2 + \lambda\int(f''(x))^2dx\right\}$$

The objective function consist of two parts, namely the residuals and the penalization of the curvature, which is the integral over the second derivative of the function. When the error becomes minimal, the function becomes very complex and therefore also very wiggly. It is therefore necessary to determine the optimal smoothing parameter λ, which determines how much the curvature of the function plays a role. Reinsch (1967) solved this problem by rewriting the function f = (f(x1), ..., f(xn)) as f = Cα, where C is cubic spline basis and α the spline coefficient. The equation can be rewritten in the following form,

$$\min_{\alpha}\|y - C\alpha\|^2 + \lambda\alpha^T H\alpha, \qquad \text{with } C\alpha = f$$

where $\|y - C\alpha\|$ is the Euclidian norm and H a penalty matrix (see Greiner 2009). The function becomes linear if $\lambda \rightarrow \infty$ and very wiggly if $\lambda \rightarrow 0$. The size of λ is found with the generalized cross validation GCV criterion, which has the following form (Hastie and Tibshirani 1990):

$$GCV(\lambda) = \sum_{t=1}^{n}\left(\frac{y_t - \hat{f}_{-t,\lambda}(x_t)}{1 - tr(H)/n}\right)^2$$

The function $\hat{f}_{-t,\lambda}(x_t)$ is the estimation of point xt, while the function f(xt) has been found with the use of all the other points xi, yi, i = 1,…, t − 1, t + 1,…, n but this one under a previously chosen λ, while tr(H) is the trace of the penalizing matrix. Minimizing the GCV iteratively using the Newton-Raphson algorithm (Greiner and Kauermann 2007), we find the optimal λ. This algorithm was implemented in R by Woods (2001).

3.2 Application of spline smoothing for the hedonic price index

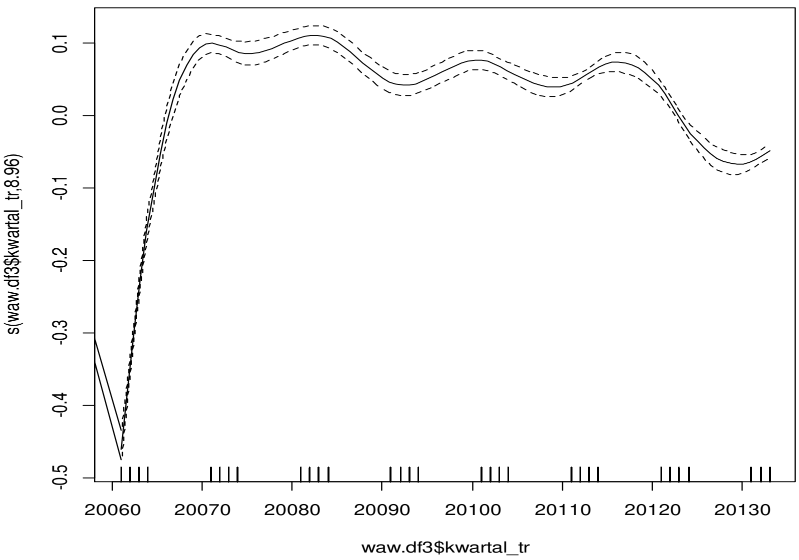
We follow Menedez et al. (2013) and use the spline method together with the GWR regression to construct the hedonic price index for Warsaw. The algorithm to combine the GWR regression and spline smoothing is based on the following loop, proposed by Menedez et al. (2013).

0. First run a usual hedonic regression with the spline function to capture the time effect ($s^t(0)$).

1. Subtract the time-spline $s^t(i)$ from our initial data, and use the de-trended data for the GWR hedonic regression $y^t(i)$.
2. Subtract the GWR results from the initial data y^t0.
3. Smooth the errors $e^t(i)$ again with the penalized spline $s^t(i)$.

Step zero initializes the procedure and allows to obtain the starting values, while the steps 1-3 are iterated until the old and new smoothed time index do not differ significantly ($\max |s^t(i) - s^{t+1}(i)| < \varepsilon,$)[6]. When the convergence criterion is reached, we have obtained the final smoothed hedonic price index, which we show in Figure 1.

Figure 1. Hedonic price index for Warsaw.



3.3 Analysis of the hedonic price index

The price index shows us how the price in a given quarter deviated from the average price for the period under consideration, controlling for the quality of housing and also its location in the city. Starting from the first quarter of 2006 to the first quarter of 2007 there was a strong price increase, which was not motivated by changes in quality or location of sold housing. This was the period were demand was booming, mainly driven by easy credit conditions, rising wages and a very high level of optimism and price expectations. Then,

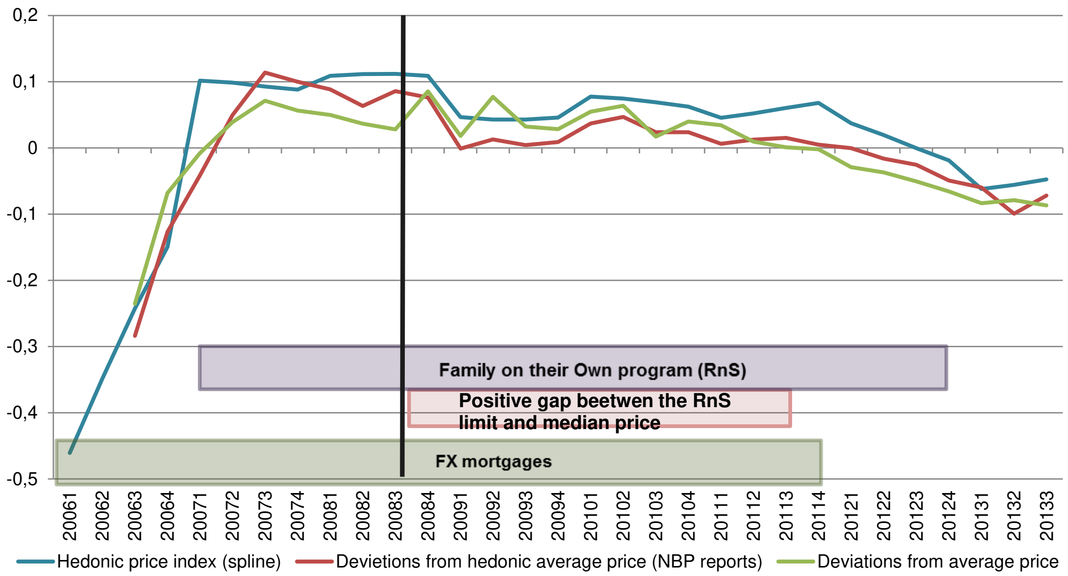
[6] The algorithm was presented by Menedez et al. (2013), however we decided to explain steps 0 and 1 in more detail in the appendix.

during 2007-2008 the price remained nearly stable and fell slightly during 2009 to 2011. Only from the beginning of 2012 and until the third quarter of that year did nominal prices show a significant decline. One of the most probable reasons was the complete cessation of FX denominated loans and the adoption of a more prudent policy of banks that were granting loans. This factor reduced housing demand and translated into falling hedonic corrected prices.

The new GWR hedonic price index as well as the old hedonic price index are better tools to analyse price changes than the mean price. The mean price is not corrected for quality changes in the analysed sample and shows deviations, which are hard to explain with economic factors. Fortunately, the hedonic price index allows us to point out the determinants of the price rises and declines in the recent period. Because the influence of policy changes is hard to pinpoint in a regression (its effects can change over time while the short time-series do not allow the application of very advanced econometric techniques that could account for it), we apply an indicator analysis (see NBP, 2013). This analysis shows that changes in income, interest rates and housing policy in the form of the Family on their Own subsidy scheme (RnS) or changes in credit regulations have a direct impact on house prices.

Firstly, we observe that the strong rise in prices in the period from 2006 Q1 to 2007 Q1 was driven by rising income and low interest rates that resulted from foreign currency denominated loans. However, at the end of 2008, when the global economic situation started to deteriorate and the Polish zloty started to depreciate, loan restrictions were imposed by banks, so the amount of available loan was reduced significantly (this moment is marked on the graph). As a direct consequence, house prices started to decline. The housing subsidy programme RnS, that was introduced in 2006 but started on a bigger scale in 2007, helped to stop the price decline. At the beginning of 2009 when the world economy started to slow down, the price limits of this programme were increased above the median price. This helped to stabilize house prices, which under other circumstances would most likely have declined. The house prices started to decline when the RnS subsidy price limits were lowered again at the end of 2011 and the programme was finally abandoned by the end of 2012. Additionally, from the beginning of 2012 banks stopped granting FX denominated loans, thus lowering housing demand and adding to the slow decline in house prices in the secondary market.

Figure 2. Important factors that influenced price movements in the secondary housing market in Warsaw.



4. Conclusions

We show in this article how not only the quality of housing, but also its location and interactions with surrounding houses affect their price. The hedonic analysis of house prices gives a better picture of the market than the analysis of mean prices. We use the geographically weighted regression to account for spatially varying estimations of regression parameters. This method, in comparison with classic OLS regression, slightly improves the results of our analysis. Applying the non-parametric estimation method on quality adjusted prices, we are able to construct a house price index, which is smooth and robust to short run price changes. This is quite a novel approach that combines the spatial and hedonic modelling of house prices, taking into account the dynamics of house values in time.

The indicator analysis of the price index shows us that the price swings were caused by an easy access to foreign currency denominated loans, and later they were kept relatively high by the Family on their own housing subsidy scheme. Only when the scheme was ended and FX loans were not granted since 2012 Q1, house prices started to decline slowly. Our analysis confirms, that new regulations or changes in existing ones and government interventions have a very big influence on house prices.

Literature

Branna J., Madej K., Będkowski M., Serdeń M., Sosiński P. and Luc M. (2012), *Analiza zależności pomiędzy ceną a lokalizacją nieruchomości na przykładzie Krakowa*, Roczniki geomatyki, X, 4(54).

Brunauer W., Feilmayr W. and Wagner K. (2012), *A new residential property price index for Austria,* Statistiken–Daten und Analysen Q3, 90-102.

Brunauer W., Lang S., Wechselberger P. and Bienert S. (2010), *Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna.* In: Journal of Real Estate Finance and Economics 41(4), 390–411.

Brunsdon C., Fotheringham A. S. and Charlton M. (1996), *Geographically weighted regression: a method for exploring spatial non-stationarity,* Geographical Analysis, 28(4), 281-298.

Cellmer, R. (2010), *Analiza przestrzenna dynamiki zmian cen nieruchomości lokalowych z wykorzystaniem regresji ważonej geograficznie*, Acta Scientarium Polonorum, Administratio Locorum, 9(4), 5-14.

Dale-Johnson, D. and Brzeski J. W. (2001), *Land value functions and land price indexes in Cracow, 1993–1999, Journal of Housing Economics*, 10(3), 307-334.

Dale-Johnson, D., Redfearn C. L. and Brzeski, W. J. (2005), *From Central Planning to Centrality: Krakow's Land Prices After Poland's Big Bang,* Real Estate Economics, 33(2), 269-297.

Fotheringham A. S., Brunsdon C. and Charlton M. (1996), *The geography of parameter space: an investigation of spatial non-stationarity,* International Journal of Geographical Information Systems, 10, 605-627.

Fotheringham A. S., Brunsdon C. and Charlton M. (2002), *Geographically Weighted Regression: the analysis of spatially varying relationships,* Wiley.

Greiner, A. (2009), *Estimating penalized spline regressions: Theory and application to economics. Applied Economics Letters* 16.18 1831-1835.

Greiner, A. and G. Kauermann, (2007). *Sustainability of US public debt: Estimating smoothing spline regressions. Economic Modelling* 24.2: 350-364.

Hastie, T. and R. Tibshirani (1990*). Generalized Additive Models*. London: Chapman and Hall.

Hastie, T. and R. Tibshirani (1993). *Varying Coefficient Models. Journal of the Royal Statistical Society*, Series B 55, 757–796.

Ligas, M., and Kulczycki, M. (2007), *Kartograficzna wizualizacja stref taksacyjnych z wykorzystaniem metod statystyki przestrzennej,* Studia i Materiały, 103.

Łaszek, J. (2013), *Housing in consumer's theory,* MPRA working paper.

Menéndez, P., Palacios, M. B., Bárcena Ruiz, M. J. and Tusell Palmer, F. J. (2011). *Measuring the Effect of the Real Estate Bubble: a House Price Index for Bilbao (No. 2011-07).* Universidad del País Vasco-Departamento de Economía Aplicada III (Econometría y Estadística).

Narodowy Bank Polski (2013). *Report on the situation in the Polish residential and commercial real estate market in 2012.*

Rosen S. (1974), *Hedonic prices and implicit markets: product differentiation in pure competition,* Journal of Political Economy 1, s. 34-55.

Reinsch, C. H. *Smoothing by spline functions*. Numerische mathematik 10.3 (1967), 177-183.

Ruppert, D., M. P. Wand and R. J. Carroll (2003). *Semiparametric Regression.* Cambridge: Cambridge University Press.

Tomczyk E. and Widłak M. (2010), *Measuring price dynamics: evidence from the Warsaw housing market*, Journal of European Real Estate Research, Vol. 3, Issue 3, 203 - 277.

Tusell F., M. B. Palacios, M. J. Bárcena and P. Menéndez (2013), *A Real-Time Property Value Index based on Web Data,* Chapter 10 in Y. Zhao and Y. Cen (Eds.). Data Mining Applications with R. Academic Press, Elsevier.

Venables, W. and B. Ripley (2003). *Modern applied statistics with S*, Volume 4. New York: Springer.

Waszczuk J. (2013), *Przestrzenne modele ekonometryczne na przykładzie warszawskiego rynku nieruchomości mieszkaniowych*, Bezpieczny Bank, 169-191.

Widłak M. and Nehrebecka N. (2011), *Zróżnicowanie cen mieszkań i ich dynamiki – ujęcie regresji kwantylowej*, Wiadomości statystyczne, Warszawa, Maj 2011.

Widłak M. (2013), *Study of factors that differentiate housing prices and the possibility of their use at NBP*, In: Report on the situation in the Polish residential and commercial real estate market in 2012, NBP.

Wood, S. (200). *Generalized additive models: an introduction with R*. CRC press.

Wood, S. N. (2001). *mgcv: Gam's and generalized ridge regression for R*. R News 1(2), 20–25.

Wood, S. N. (2014). *The mgcv Package*. The Comprehensive R Archive Network.

Appendix:

1. **Description of the explanatory variables. The variables that serve as the benchmark for dummy regressions are underlined.**

stand_1: dummy variable taking 1 if the flat is of high standard

<u>*stand_2:*</u> dummy variable taking 1 if the flat is of medium standard

stand_3_4: dummy variable taking 1 if the flat is of low standard or even unfinished

<u>*r_1_2*</u>: dummy variable taking 1 if the flat was constructed before 1944

r_3: dummy variable taking 1 if the flat was constructed during 1945-1970

r_4: dummy variable taking 1 if the flat was constructed during 1971-1978

r_5: dummy variable taking 1 if the flat was constructed during 1979-1988

r_6: dummy variable taking 1 if the flat was constructed during 1989-2001

r_7: dummy variable taking 1 if the flat was constructed during 2002-2005

r_8: dummy variable taking 1 if the flat was constructed after 2005

rok_nowe: dummy variable taking 1 if the flat was constructed at most one year before it was sold (it accounts for flats that were still under construction)

sq_pow: total area in sq. m, squared

ilepokoi_1_2: dummy variable taking 1 if the numer of rooms is 1 or 2

liczba pokoi: numer of rooms in the flat

powierzchnia: total area in sq. m

l_stacja_m: logarithm of euklidian distance to the nearest metro station

parki_L1000: number of parks in a distance up to 1000 m

l_centrum_dist: logarithm of euklidian distance to the city centre of Warsaw (Pałac Kultury i Nauki)

zielen_1: dummy variable taking 1 if the flat is nearer than 350 m away from a green area

mokotow: dummy variable taking 1 if the flat is located in the city quarter Mokotów

ochota: dummy variable taking 1 if the flat is located in the city quarter Ochota

<u>*śródmieście:*</u> dummy variable taking 1 if the flat is located in the city quarter Śródmieście

targówek: dummy variable taking 1 if the flat is located in the city quarter Targówek

ursynów: dummy variable taking 1 if the flat is located in the city quarter Ursynów

wilanów: dummy variable taking 1 if the flat is located in the city quarter Wilanów *wola:* dummy variable taking 1 if the flat is located in the city quarter Wola

żoliborz: dummy variable taking 1 if the flat is located in the city quarter Żoliborz

gr1: dummy variable taking 1 if the flat is located in the city quarter Bielany or Bemowo

gr2: dummy variable taking 1 if the flat is located in the city quarter Białołęka or Targówek

gr3: dummy variable taking 1 if the flat is located in the city quarter Praga Północ or Praga Południe

gr4: dummy variable taking 1 if the flat is located in the city quarter Wawer or Rembertów

gr5: dummy variable taking 1 if the flat is located in the city quarter Ursus Włochy, 0 wpp

<u>t_20061</u> - t_20133: dummy variable taking 1 if the flat was sold in a given year and quarter

2. Spatial distribution of residuals:

Figure 3. Spatial distribution of residuals, OLS model with spatial variables

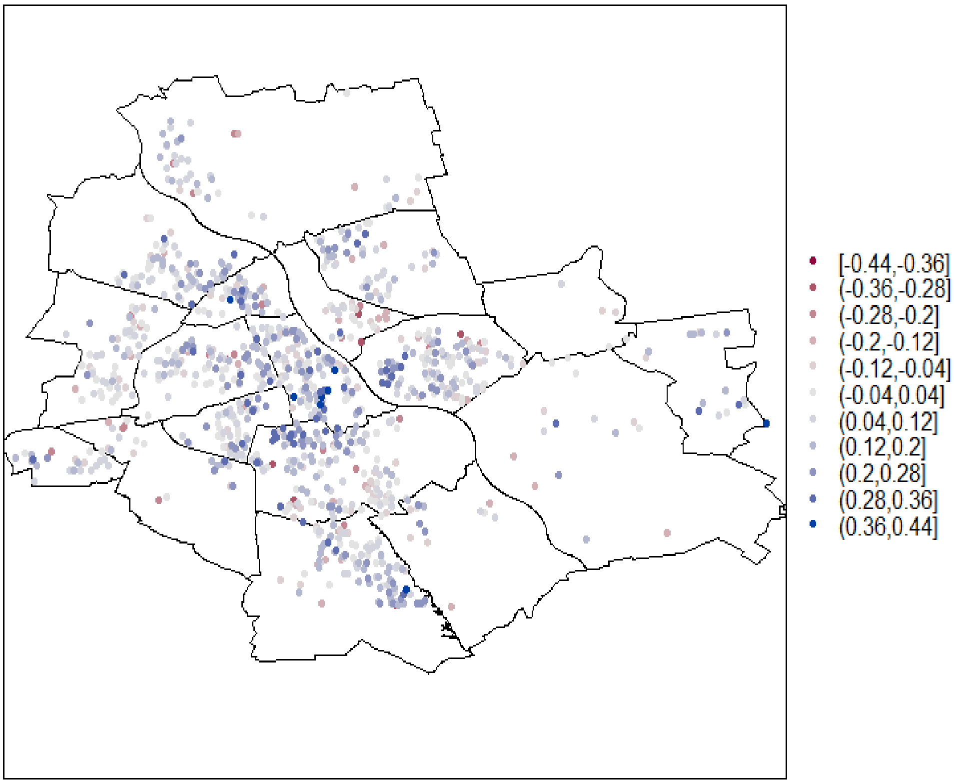
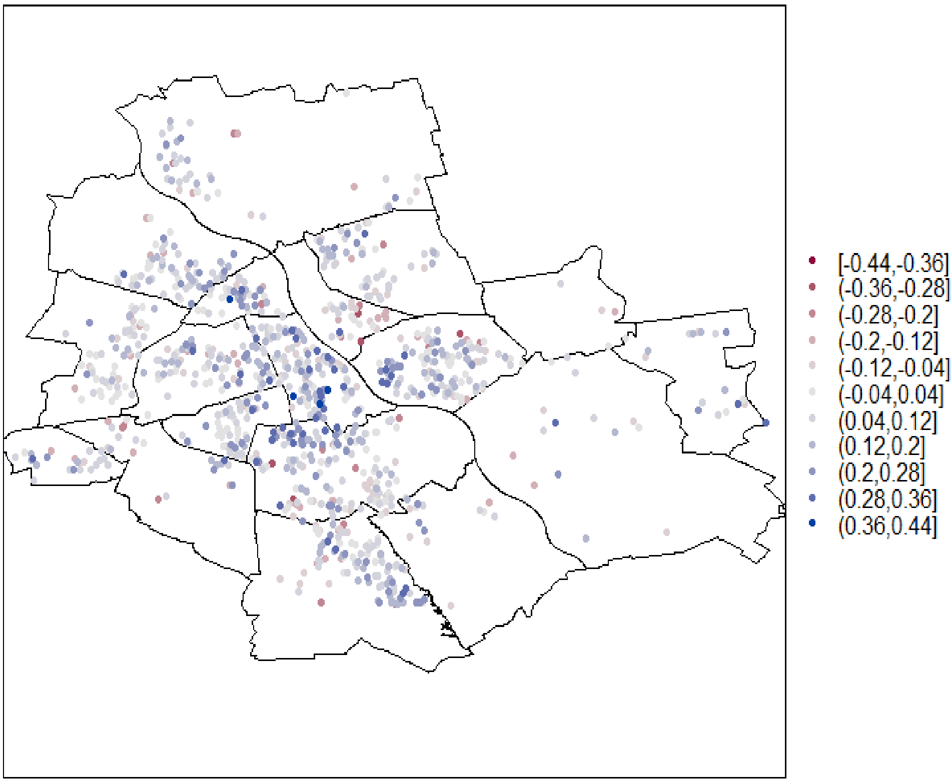


Figure 4. Spatial distribution of residuals, GWR



3. Explanation of the backfitting algorithm that combines the GWR and smooth spline regression

Step 0: Create the initial spline model *gam1.mod* from pure data, where our dataset is called *wawa.df1*.

```
gam1.mod <- gam(lcena_m2 ~ l_stacja_m + zielen_1 + l_centrum_ + powierzchn + sq_pow + ilepokoi_1 + stand_1 + stand_3_4 + spoldzielc + rok_1_2 + rok_3 + rok_4 + rok_5 + rok_7 + rok_8 + rok_nowe + gr1 + gr2 + gr3 + gr4 + gr5 + mokotow + ochota + wola + ursynow + zoliborz + wilanow + s(kwartal), data=wawa.df1)
```

and extract the value of the spline with the help of the *terms* option.

```
pred = predict(gam1.mod,wawa.df1, type="terms")
```

finally, we save the predicted spline values as waw.df2

```
waw.df2=as.data.frame(pred)
```

and combine the new database with the initial transaction price and information about the transaction quarter and call it *waw.df3*

```
waw.df3 = cbind(waw.df2, lcena_m2, kwartal_tr)
```

It is important to rename the value of the spline so that R treats it as a number and not a function. Remember that so far the value of the spline was a whole function, thus we cannot perform any usual calculations. We change the name s(kwartal) to s_kwartal, and R treats this variable as a number, and finally subtracts the spline value from the initial transaction price.

```
waw.df3$bezspline <- waw.df3$lcena_m2-waw.df3$s_kwartal
```

From here, we can simply follow the procedure proposed by Menedez et al. (2013), thus we run the GWR regression on the de-trended data, apply the spline smoothing to the residuals, de-trend the data again, run the GWR etc., until we cannot improve the spline $\max|s^t(i) - s^{t+1}(i)|<\varepsilon$.