Analysis of Spatial Effects in House Prices

Erdi Kızılkaya
Specialist
Statistics Department, Central Bank of the Republic of Turkey
www.tcmb.gov.tr
Ankara, Turkey
erdi.kizilkaya@tcmb.gov.tr

Abstract

Spatial interaction is that closer units are more correlated than distant ones and have similar characteristics. This study explores the potential spatial interaction in house prices in Ankara and aims to model a housing price index that includes the interaction in question. We employ Moran's I test statistic to test the statistical significance of the spatial dependence and we form spatial weight matrices for seven districts of Ankara that have sufficient data. We explain the selection procedure of appropriate spatial econometric models in detail and compare changes in coefficients of independent variables calculated by Spatial Error Model (SEM) to the ones calculated by the Ordinary Least Squares (OLS) method. We conclude that SEM offers better fit since the AIC values of these models are lower compared to OLS. We construct indices for each district separately and we introduce a new house price index for Ankara by weighting district-based indices. We compare the newly constructed Spatial Hedonic House Price Index with the house price indices already published by the Central Bank of the Republic of Turkey, which are quality adjusted Hedonic House Price Index and House Price Index calculated by stratified median price method. Results show that spatial analysis enables obtaining more reliable results when data set contains spatially correlated observations.

1. Introduction

Since housing investments have a significant share within the total wealth of households and a notable portion of the credit stock comprises of the housing loans, it is critical to construct a reliable house price index for institutions that pursue financial stability. However, this is a difficult task due to the heterogeneous nature of the housing market. One of the most common methods used to measure house prices is to calculate the mean and median prices of the houses sold but this method may produce misleading results because it does not take into account quality changes. Hedonic methods enable measuring quality changes by including the characteristics of the dwellings such as size of the house, number of rooms and number of bathrooms in the calculation. However, location is a fundamental determinant in the housing market and in order to construct a robust indicator it is very important to specify a hedonic model that takes into account the spatial interaction among house prices (Syed et al., 2008).

Although it is a crucial indicator in monitoring financial stability, there was no official house price index in many countries, including Turkey until the recent global financial crisis. In order to fill this data gap, Central Bank of the Republic of Turkey (CBRT) began to publish the House Price Index (HPI) calculated using stratified median price method in 2012. Nominal price change measured by HPI contains both change in the quality and in the price of the dwellings. Housing market is inherently heterogeneous, thus two different groups of dwellings may have different characteristics. This may create a bias on the HPI, so for accurate measurement, it is necessary to measure the impact of quality changes on the prices. Since the increase in house prices includes both pure price changes and quality changes, a price increase due to quality change can be considered as a bubble in a misleading way (Hülagü et al., 2016). Hedonic regression methods adjust for quality change, hence allow measuring the pure price change (Eurostat Handbook, 2013). In order to identify the source of the price increase in Turkish housing market correctly, CBRT began to publish Hedonic House Price Index (HHPI) in 2016. Commonly used hedonic regression models do not take into account the spatial interaction of house prices (Rambaldi and Rao, 2011). In other words, traditional regression models ignore the spatial effect by assuming that observations are independent. According to Waldo Tobler's first rule of geography (1970), everything is related to everything else but near things are more related compared to distant ones. In the light of this information, we can say that while constructing house price indicators, including analysis of spatial interactions besides hedonic price methods that measure pure price change will produce more reliable results. In this regard, we aim to construct a more sophisticated index by taking into account the effect of spatial interactions and hence to advance the HHPI studies.

This study explores the potential spatial interaction in house prices in Ankara, capital city of Turkey, and aims to model a house price index that includes the interaction in question. First, we examine the necessity of spatial econometrics and how it contributes to the current house price index studies. Then, we analyze whether house prices in large districts of Ankara have spatial interaction. We explain the weight matrix and Moran's I statistic employed to test the significance of spatial interaction. We describe the selection procedure of the spatial econometric model (Spatial Error Model or Spatial Lag Model) performed according to the analysis results. We compare results of spatial econometric model with the results of ordinary least squares (OLS) model and examine the changes in coefficients of independent variables. Finally, we construct a Spatial Hedonic House Price Index, which includes spatial effects in the calculation, and we compare the new index with the HPI and HHPI published by the CBRT.

2. Spatial Econometrics

Spatial econometrics is a subfield of econometrics that deals with spatial interaction and spatial structure (Anselin, 1988). It extends traditional econometric methods by taking into account the potential impact of location of data in question (LeSage and Pace, 2009). For this reason, spatial analysis is based on the assumption that an event is related to the events taking place in neighboring locations, in other words, that the location is crucial. Case (1991) states that spatial modeling has a wide range of use. Taxes, levels of expenditure, unemployment, poverty and firm behavior are examples of the areas of spatial analysis. According to Ertur and Koch (2007), countries cannot be considered spatially independent observations and growth models should include spatial interactions because of technological interdependence. Another analysis performed by spatial modelling is from the field of criminology and finds out that homicide is strongly clustered in space in the US from 1960 to 1990 (Baller et al.,2001). In their study, Rey and Montouri (1999) employ spatial analysis to examine the question of regional income convergence in the US states and prove that there is a strong spatial autocorrelation in regional income growth patterns. In the study examining the housing price movements in the US, a significant spatial effect is identified among the neighboring states (Holly et al., 2010).

The need to include the effects of differences in location into the statistical analysis necessitates spatial econometrics. The two main types of spatial regression are Spatial Lag Model (SLM) and Spatial Error Model (SEM). SLM is appropriate in the case of spatial dependence and examines the correlation of the dependent variable with the neighboring dependent variables. SEM, on the other hand, is appropriate in case of spatial heterogeneity and examines the spatial dependence of error terms that may occur due to unobserved variables or shocks. Presence of spatial autocorrelation or heteroscedasticity violates OLS assumption of independent and identically distributed variables. Spatial dependence may result in biased and inefficient OLS estimates, and hence misleading deductions. Therefore, potential spatial interaction among variables or error terms should be considered when working with spatial data. Conceptual comparison of three aforementioned models, OLS, SLM and SEM, in terms of spatial dependence among variables and error terms is summarized in Figure 1 (Whiteacre, 2017).

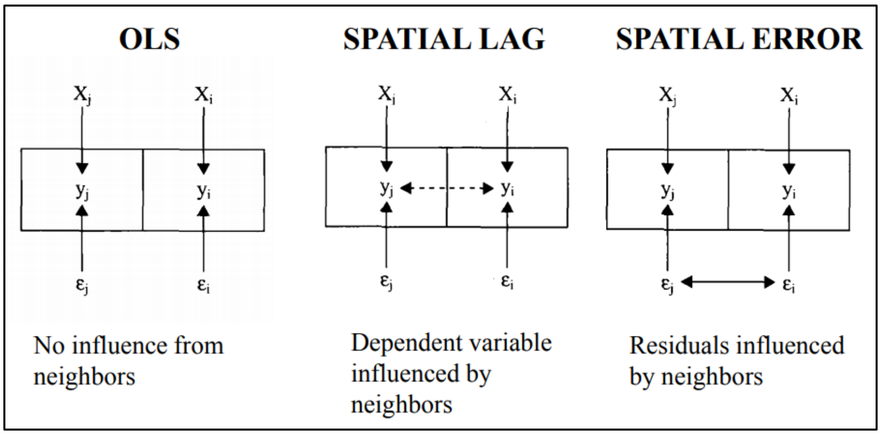


Figure 1. Model comparison

Whether spatial regression is necessary and when necessary which model should be selected is demonstrated by Anselin (2005) employing a decision tree (Figure 2). At the first stage, OLS regression is run and spatial dependence is tested using “Lagrange Multiplier” tests. If spatial dependency is not statistically significant, then OLS results are used, otherwise the model is specified according to whether the spatial dependence is in error terms or dependent variable.

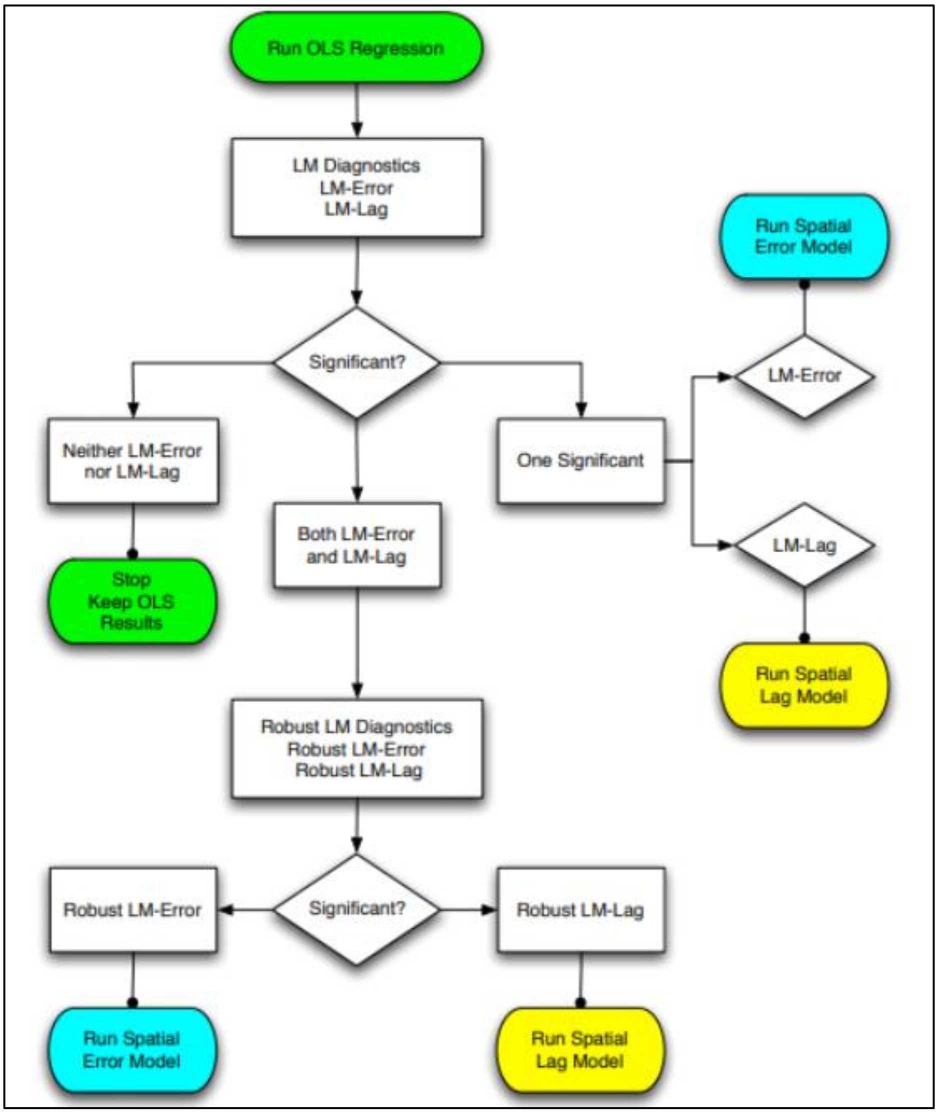


Figure 2. Spatial model decision tree

"Moran's I Statistic" developed by Patrick Alfred Pierce Moran (1950) is employed to test the statistical significance of spatial dependence. Briefly, Moran's I statistic measures the correlation between the mean values of neighboring observations (Ward and Gleditsch, 2007). Moran's I values range from -1, indicating a perfect dispersion, to 1, indicating a perfect clustering. Value of 0 represents a random spatial pattern. (Figure 3) Since these values (-1, 0, 1) are not observed in many cases, the statistical significance of Moran's I statistic is tested and when it is significant, it is concluded that spatial autocorrelation is present.

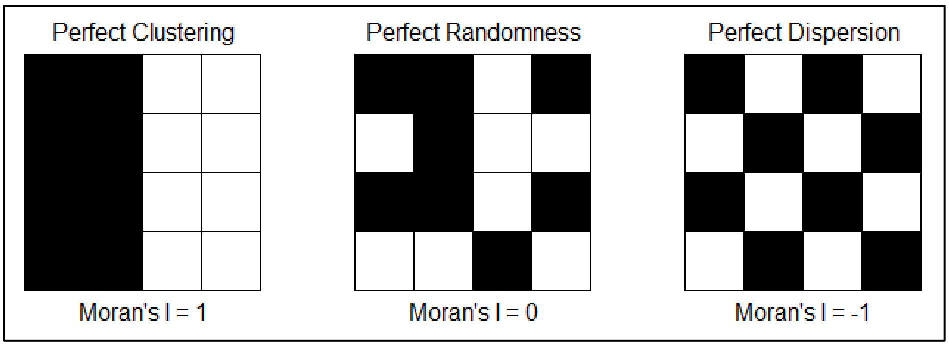


Figure 3. Examples of Moran's I

Specifying the weight matrix is a critical step in spatial econometric analysis because weight matrix represents the spatial structure of the data and enables measuring spatial correlation. The spatial weight matrix provides information on which regions are neighbors and it can be formulated either by boundaries of the units or by distance between the units.

3. Data

Data source is the valuation reports prepared by real estate appraisal companies at the time of approval of individual housing loans. Values at these reports are used as a proxy for price since actual sales price is not available. The actual sale of the property and utilization of the loan is not required and all houses appraised are included in the scope. Our data set contains over three million observations from all over Turkey between 2010 and 2017. As mentioned above our analysis focuses on observations from Ankara which comprises 13% of total data set. Ankara has the second largest share of data among all cities, following Istanbul (Figure 4).

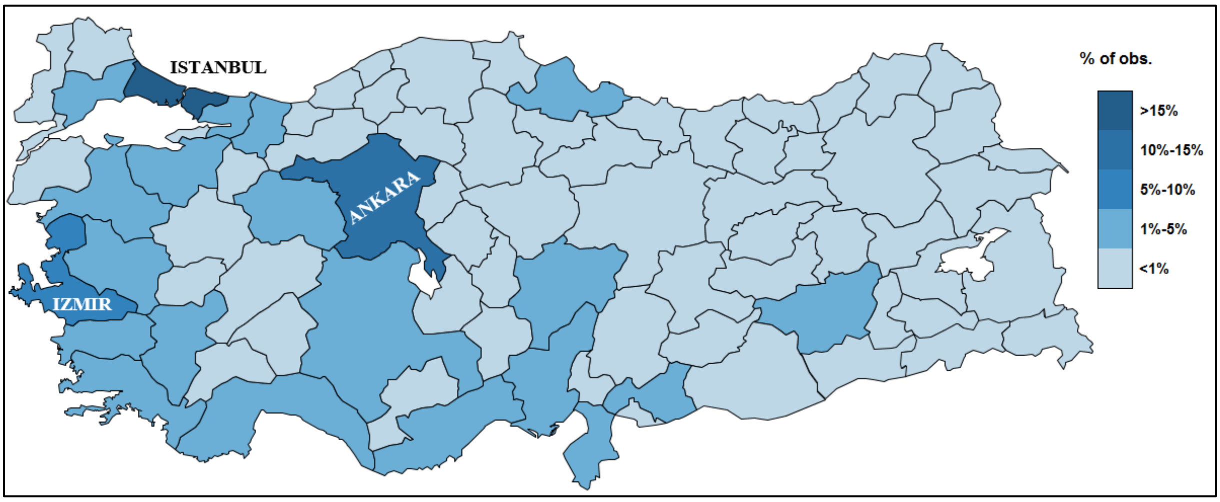


Figure 4. Data Distribution Map of Turkey

Data set consists of about 40 variables regarding the address information as well as characteristics of dwellings appraised such as gross area of use, number of rooms or heating type. In order to calculate a house price index containing the spatial effects, geographic location of the houses in the data set is required. In other words, latitude and longitude information of the houses is necessary to measure the spatial interaction statistically. However, geocodes of the dwellings have been missing in the data set until 2018. Therefore, Sharing System Protocol was signed with Land Registry and Cadaster and geocodes have been provided through a web service tool where we match address information with latitude and longitude of the dwellings. Since the address information of the dwellings is in text format, there are many misspellings or missing data in the data set. This resulted in partial matchup with Land Registry database and latitude and longitude information of some of the dwellings could not be delivered. Table 1 shows the matchup ratio of observations for the largest districts of Ankara.

Table 1. Number of Observations (Districts of Ankara)

Year	Keçiören	Mamak	Etimesgut	Çankaya	Yenimahalle	Sincan	Others	Total
2010	7.966	4.014	6.758	5.971	3.939	1.806	4.467	34.921
2011	8.064	4.274	6.510	5.926	4.463	2.035	4.953	36.225
2012	6.582	3.394	5.197	4.590	3.719	1.416	4.480	29.378
2013	9.996	5.356	7.483	6.159	5.672	2.177	6.732	43.575
2014	8.961	5.305	6.651	4.953	4.262	2.855	6.767	39.754
2015	9.737	6.518	6.906	5.044	4.647	2.490	8.275	43.617
2016	10.445	6.943	6.728	5.010	4.810	2.890	9.577	46.403
2017	10.233	7.355	6.600	5.375	5.080	3.504	10.563	48.710
Total	**71.984**	**43.159**	**52.833**	**43.028**	**36.592**	**19.173**	**55.814**	**322.583**
Matchup Ratio	**89,0%**	**67,8%**	**85,8%**	**77,2%**	**71,6%**	**38,7%**	**76,5%**	**74,1%**

4. Analysis and Application

We conduct spatial analysis for 7 large districts of Ankara namely, Altındağ, Çankaya, Etimesgut, Keçiören, Mamak, Sincan and Yenimahalle, where there are sufficient number of observations. At first, we set up the model for each district based on the regression models and variables used in the HHPI study.[1] Dependent and independent variables are summarized in Table 2.

Table 2. Model Variables

Variable	**Description**
Log Price (D.V.)	Logarithm of appraisal value
Gross Area of Use	Total area of dwelling
Quality of Construction	Higher quality=1 Lower Quality=0
No of Rooms	Total number of rooms
No of Bathrooms	Total number of bathrooms
No of Balconies	Total number of balconies
Security Service	Gated community=1 No gated community=0
Elevator	Available=1 Not available=0
Year of Construction	Construction year of dwelling
Heating	Central heating or wall hung gas boiler system=1 Other=0

In particular, our log-linear regression model is as follows:

$$lnp_n^t = \beta_0^t + \sum_k \beta_k^t z_{nk}^t + \varepsilon_n^t \qquad \forall n, t \tag{1}$$

where p_n^t is the price of the property, z_{nk}^t is the characteristic k of the property, β_k^t is shadow price of the features of the property and ε_n^t is the error term.

[1] For detailed information: Hülagü, T., Kızılkaya, E., Özbekler, A. G., & Tunar, P. (2016). A hedonic house price index for Turkey. Central Bank of the Republic of Turkey Working Papers, (16/03).

However, when we analyze the matched data we find out that "year of construction" and "heating type" variables are no longer statistically significant. Heating type is a dummy variable, which takes "0", or "1" and when we examine the data set by districts, we see that the average value of this variable is between 0.93-0.95 during the analysis period. We think that this variable is not statistically significant since there is not enough variation in the new data set. Selected variables for each district are shown in Table 3.

Table 3. Variable List

	Districts of Ankara						
Variables	**Altındağ**	**Çankaya**	**Etimesgut**	**Keçiören**	**Mamak**	**Sincan**	**Yenimahalle**
Gross Area of Use	✓	✓	✓	✓	✓	✓	✓
Quality of Construction	✓	✓	✓	✓	✓	✓	✓
No of Rooms	✓	✓	✓	✓	✓	✓	✓
No of Bathrooms	✓	✓	✓	✓	✓	✓	✓
No of Balconies	✓	✓	✓	✓	✓	✓	✗
Security Service	✗	✓	✓	✓	✓	✗	✓
Elevator	✓	✓	✓	✓	✓	✓	✓
Year of Construction	✗	✗	✗	✗	✗	✗	✗
Heating	✗	✗	✗	✗	✗	✗	✗
✓=included in regression model, ✗=not included							

In the first phase of the study, we form weight matrices by distance-based method in 96 months for each of the 7 districts. Then, we test the spatial correlation of the regression results calculated by the OLS method with Moran's I statistic. After identifying the spatial dependence, we determine the spatial model to be employed using the "Spatial Regression Decision Tree" (Figure 2).

According to the results of Moran's I test statistics, we find out a statistically significant spatial dependence for all districts in all periods. Figure 5 is an example of test results in Yenimahalle district for January 2012. Test results show that independent error terms assumption is violated, hence we conclude that spatial models should be applied. In order to choose the spatial model to be applied, we employ "Robust LM (lag)" and "Robust LM (error)" tests. For some periods, Robust LM (lag) test is not statistically significant and for others where both tests are significant, Robust LM (error) test takes higher values. In the light of this information, we decide to apply Spatial Error Model as showed in formula (2) and (3) in constructing the Spatial House Price Index for Ankara.

$$Y = X\beta + \varepsilon \tag{2}$$

where Y is an Nx1 vector of observations on the dependent variable, X is an NxK matrix of observations on the independent variables, β is a Kx1 vector of regression coefficients and ε is an Nx1 vector of spatially autocorrelated error terms.

$$\varepsilon = \lambda W_{\varepsilon} + u \tag{3}$$

where W_{ε} is a spatial lag for the errors, λ is the autoregressive coefficient and u is another error term.

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.1835	19.7644	0.00000
Lagrange Multiplier (lag)	1	154.9263	0.00000
Robust LM (lag)	1	37.5714	0.00000
Lagrange Multiplier (error)	1	279.0973	0.00000
Robust LM (error)	1	161.7425	0.00000
Lagrange Multiplier (SARMA)	2	316.6687	0.00000

Figure 5. Spatial dependency test results

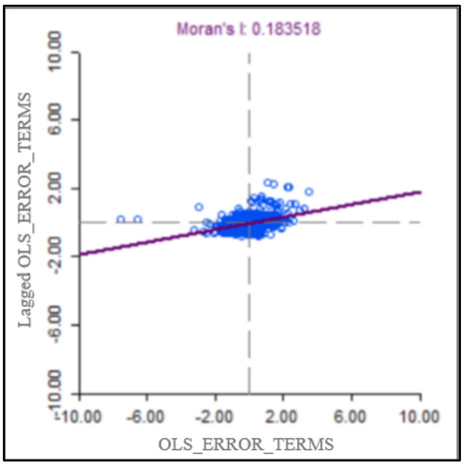


Figure 6. Moran's I test statistics for Yenimahalle district

At the next stage of the study, we apply OLS and SEM methods for all the districts in question for each period. We examine the effects of the selected independent variables on the dependent variable, namely price of the dwelling. Since we detect spatial dependence in all districts for each period, we assume that use of spatial models will provide more reliable results. We compare the results of OLS and SEM methods in order to investigate the validity of this assumption.

When we examine the analysis results for the largest district, Çankaya, we see that spatial error dependence coefficient (λ) is statistically significant for SEM models (Table 4). Security service increases the price of a dwelling by 36.5% according to the OLS model whereas 28.9% according to the SEM. This difference may stem from the fact that OLS models have the problem of overestimated coefficients when there is spatial correlation within the data set. Having a good or high quality construction increases the price of a dwelling by 11.8% in OLS method and by 13.3% in SEM method. The coefficients of the elevator, gross area of use, number of rooms, number of balconies and number of bathrooms are very close in both methods. The AIC value of the model estimated by SEM method is lower, hence we can say that SEM method provides better fit.

Table 4. Results of Çankaya (July 2011)

	Dependent Variable: Log(Price)	
	Coefficients	
Variables	**OLS**	**SEM**
Constant	10.960***	11.075***
Security Service	0.365***	0.289***
Elevator	0.131***	0.138***
No of Rooms	0.112***	0.114***
No of Bathrooms	0.068***	0.062***
No of Balconies	0.061***	0.066***
Gross Area of Use	0.003***	0.003***
Quality of Construction	0.118***	0.133***
Lambda (λ)		0.416***
Akaike info criterion (AIC)	959.894	707.126
No of Obs.	1598	
***,**,* denotes significance at 1%, 5% and 10% level, respectively.		

As explained above, for the 96 months between 2010 and 2017, we employ models generated by SEM method for each of the 7 districts. In other words, we calculate 96 regression models for each district with the aim of producing a house price index which takes into account spatial dependence as well as quality effect. We use the mean values of the independent variables (characteristics of dwellings) in the designated base period and obtain the coefficients using SEM method. We calculate indices by specifying January 2012 as base period with the formula below for 7 districts.

$$P_i^t = \frac{exp(\widehat{\beta_0^t})exp\left[\sum_k \widehat{\beta_k^t}\overline{z_{nk}^0}\right]}{exp(\widehat{\beta_0^0})exp\left[\sum_k \widehat{\beta_k^0}\overline{z_{nk}^0}\right]} \tag{4}$$

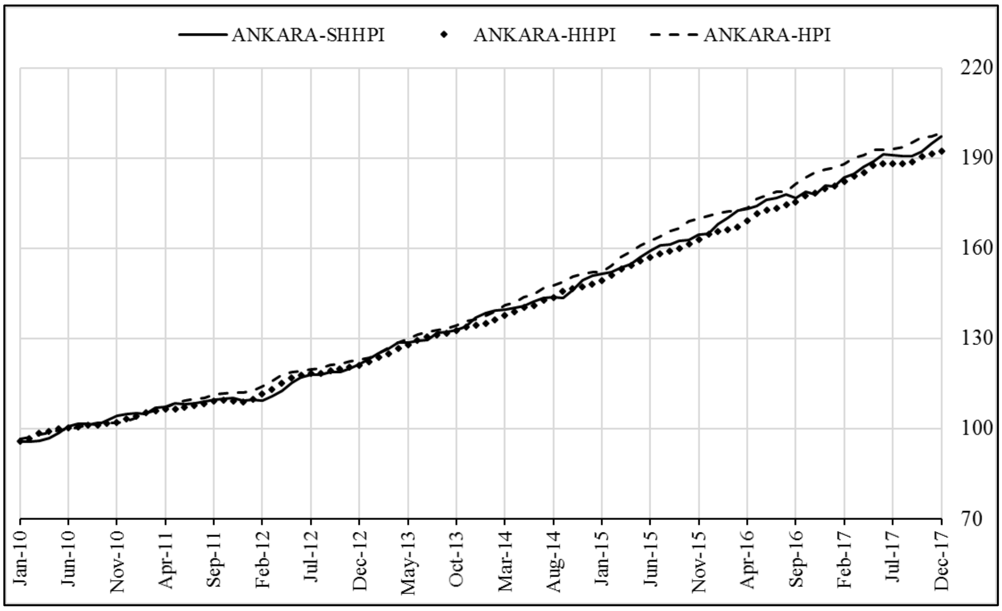
P_i^t: house price index for stratum i in period t
$\widehat{\beta_k^0}$: estimated shadow price of the features in the base period
$\widehat{\beta_k^t}$: estimated shadow price of the features in period t
$\overline{z_{nk}^0}$: average housing characteristics in the base period

After calculating the indices for 7 districts, house price index for Ankara is generated by weighting the district indices. We use number of houses sold the previous year as weight.

$$P^t = \sum w_i^T * p_i^t \tag{5}$$

P^t: house price index for Ankara in period t
w_i^T: weight for stratum i in year T
p_i^t: house price index for stratum i in period t

In order to be able to compare the results with HPI and HHPI, we rebase the index to 2010=100 and we name the index Spatial Hedonic House Price Index (SHHPI). All three indices are shown in Graph 1.



Graph 1. Comparison of House Price Indices of Ankara

5. Results and Conclusion

SHHPI for Ankara is calculated as 197.21 in December 2017. In the same month, HHPI takes the value of 192.28 and HPI takes the value of 198.37. According to the results, in the period between 2010 and 2017 HPI has increased by 98.37% and showed the highest increase among three indices. In the same period, SHHPI and HHPI have increased by 97.21% and 92.28%, respectively. Since the rate of increase of consumer price index is 83.53% (83.24% for Ankara) in the same period we can say that all indices have increased in real terms, too. Accordingly, the real increase rates between 2010 and 2017 are 7.45% for SHHPI, 4.77% for HHPI and 8.09% for HPI.

According to the findings, the new index, namely SHHPI, shows a similar pattern with HHPI and has a slightly lower level than HPI. However, at some points SHHPI approaches HPI but later converges back to HHPI. We can say that SHHPI is more volatile compared to other indices. We can conclude that in order to avoid the problem of overestimation of coefficients when interpreting the effect of independent variables on house prices, SEM method provides better fit compared to OLS method. Results prove that spatial analysis enables obtaining more reliable outcomes when data set contains spatially correlated observations.

References

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Anselin, L. (2005). Exploring Spatial Data with GeoDa^{TM}: A Workbook p.199. Retrieved from http://www.csiss.org/clearinghouse/GeoDa/geodaworkbook.pdf.

Baller, R.D., Anselin, L., Messner, S.F., Deane, G. ve Darnell, F.H. (2001). Structural Covariates of U.S. County Homicide Rates: Incorporating Spatial Effects. Retrieved from https://onlinelibrary.wiley.com/doi/full/10.1111/j.1745-9125.2001.tb00933.x.

Case, A. (1991). Spatial Patterns in Household Demand. Econometrica, 59(4), 953-965. doi:10.2307/2938168

Ertur, C., & Koch, W. (2007). Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence. Journal of Applied Econometrics, 22(6), 1033-1062. Retrieved from http://www.jstor.org/stable/25146563

Eurostat Handbook (2013). European Commission, Eurostat, Products Manuals and Guidelines. Retrieved from, https://ec.europa.eu/eurostat/en/web/products-manuals-and-guidelines/-/KS-RA-12-022

Holly, Sean & Pesaran, Hashem & Yamagata, Takashi. (2010). A Spatio-Temporal Model of House Prices in the US. Journal of Econometrics. 158. 160-173. 10.1016/j.jeconom.2010.03.040.

Hülagü, T., Kızılkaya, E., Özbekler, A.G. ve Tunar, P. (2016). A Hedonic House Price Index for Turkey CBRT Working Papers. Retrieved from http://www.tcmb.gov.tr/wps/wcm/connect/EN/TCMB+EN/Main+Menu/Publications/Research/Working+Paperss/2016/16-03

Rambaldi, Alicia & Rao, D.S.. (2011). Hedonic Predicted House Price Indices Using Time-Varying Hedonic Models with Spatial Autocorrelation.

Rey, Sergio and Montouri, Brett, (1999), US Regional Income Convergence: A Spatial Econometric Perspective, Regional Studies, 33, issue 2, p. 143-156.

Syed, Iqbal & Hill, Robert & Melser, Daniel. (2008). Flexible Spatial and Temporal Hedonic Price Indexes for Housing in the Presence of Missing Data. SSRN Electronic Journal. 10.2139/ssrn.1313693.

Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography, 46, 234-240. doi:10.2307/143141

Ward, M. D., ve Gleditsch, K. S. (2008). Quantitative Applications in the Social Sciences: Spatial regression models. Thousand Oaks, CA: SAGE Publications Ltd.

Whiteacre, B. (2017). Spatial Econometrics Workshop, Western Agricultural Economics Association Annual Meeting. Retrieved from http://agecon.okstate.edu/faculty/publications/5538.pdf.