SPATIAL DEPENDENCY AND HETEROGENEITY IN HOUSING PRICES ACROSS UKRAINIAN CITIES

by

Valentyna Katsalap

A thesis submitted in partial fulfillment of
the requirements for the degree of

Master of Arts in Economics

National University "Kyiv-Mohyla Academy"
Master's Program in Economics

2008

Approved by ____
Mr. Volodymyr Sidenko (Head of the State Examination Committee)

Program Authorized
to Offer Degree Master's Program in Economics, NaUKMA

____

Date ____

National University "Kyiv-Mohyla Academy"

Abstract

SPATIAL DEPENDENCY AND HETEROGENEITY IN HOUSING PRICES ACROSS UKRAINIAN CITIES

by Valentyna Katsalap

Head of the State Examination Committee: Mr. Volodymyr Sidenko,
Senior Economist
Institute of Economy and Forecasting,
National Academy of Sciences of Ukraine

Apartments are heterogeneous goods, which are valued not only according to their structural characteristics (number of rooms, total living area, repair), but also by location and neighbor characteristics (unemployment rate, ecological situation, quality of schools etc.). Similar apartments located in different places are valued differently and spatial heterogeneity is observed. At the same time prices of apartments located in neighbor cities are correlated. In this paper I apply spatial analysis to eight Ukrainian regional centers and found that housing prices in neighbor cities directly affect each other, while omitted neighbor variables lead to the spatial correlation in error terms. This research can be used by real estate experts, home sellers and buyers for making decisions and forecasting future prices in some city after a change in prices in neighbor cities was observed.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

First of all, I would like to express my gratitude to my thesis advisor Tom Coupe for reading and correcting all my drafts, giving valuable advises and comments and encouraging me.

I would like to express my special thanks to Oleksandr Shepotylo for his course in spatial econometrics, which was very helpful for the estimation part of my thesis.

Research workshop professors gave many useful comments; therefore, I am grateful to Olesya Verchenko, Andriy Zapechelnyuk, Pavlo Prokopovych, Ganna Vakhitova.

This list would not be complete if I didn't mention here my friends. Scores of times I discussed my thesis with Irynka Hayduk, Serhiy Pysarenko, Roman Voznyak and Oleksandr Kubatko, who patiently listened to my complaints, sometimes read my drafts, gave advices and always expressed their belief in my ability to write a good thesis.

GLOSSARY

Spatial heterogeneity – structural instability, which can be represented by varying coefficients or in the form of non-constant variance (heteroskedasticity).

Spatial dependency – neighbor values are correlated.

Spatial lag – a weighted average of neighbor values.

Spatial weights matrix – an $n \times n$ matrix which assigns positive weights to neighbor values and zero to non-neighbor values.

.

Chapter 1

INTRODUCTION

"Everything is related to everything else, but near things are more related than distant things."
(First law of geography by Waldo Tobler)

The introduction of the principles of geography into economics has produced a field called spatial economics. The major point in spatial economics is that not only characteristics of the object have value, but also its location. Spatial economics is especially applicable to the housing market, since a house has a fixed location. Consequently, houses located in different places may be valued differently, but the valuation of the houses in the neighbor cities could be correlated. The latter feature is called spatial dependency and the former – spatial heterogeneity.

Besner (2002) gives an argument in favor of spatial dependency – prices of houses in neighbor cities are used as a benchmark for home buyers and sellers, especially for very similar houses (similarity in terms of living area has the highest weight), while Karlsson (2007) explains spatial heterogeneity as result of individual's preferences for access and amenity values.

There are many models that take into account spatial aspects of housing price analysis, but there is no consensus which model has the most accurate prediction and the highest explanatory power. Though most of the studies report that spatial hedonic models they used performed better than OLS (Bitter et al. (2006) for Tucson (Arizona), Long et al. (2007) for Toronto, Besner (2002) using for

Montreal), there are also opposite findings – spatial models did not outperform traditional hedonic price model for Tokyo (Gao et al. (2002))[1].

Housing prices in Ukraine's largest cities are widely discussed in the press due to their persistent rise and according to estimates of "Expert" magazine estimates if a family lived off subsistence consumption, it would take 21 years to afford a one-bedroom apartment with total area 45 sq. meters for a family in Kyiv or Kharkiv, 49 years in Odesa, 33 years in Lviv, 26 years in Dnipropetrovsk, 20 years in Cherkasy, 23 years in Rivne and 41 years in Zhytomyr. This information reflects the relationship between average income and housing prices in some oblast cities of Ukraine and it creates an impression that there is significant difference in housing prices across Ukrainian cities, which can be confirmed by estimates of real estate experts represented in Table 1.

Table 1.1: Price per square meter in oblast cities with different population (ranked from the highest to the lowest)[2]

City	Price per sq. meter, $	Population, thousand
Kyiv	2600	2708
Odesa	1820	1002
Dnipropetrovsk	1340	1047
Kharkiv	1170	1463
Lviv	1100	735
Rivne	925	248
Zhytomyr	920	276
Cherkasy	820	292

[1] Different spatial models (kriging, spatial autoregressive models, spatial error model, geographically weighted regression) and their performance are discussed in the next section in more details.

[2] Estimates of Association of Realtors of Ukraine (first half of 2007)

The natural question is whether housing attributes (such as repair, type of material of which the house is made) are valued the same in different places (spatial heterogeneity) and if housing prices are interrelated in Ukrainian cities (spatial dependency). And what matters more - geography (cities located closer geographically have correlated housing prices) or socio-economic characteristics (cities similar in terms of population have correlated prices). These hypotheses will be checked in this thesis with the data on housing prices in eight Ukrainian oblast centers (which is roughly one third of the total number of oblast centers) and in districts of Kyiv using spatial autoregressive model, spatial error model and spatially autoregressive model with spatially correlated error terms. It is the first attempt to explore the relationship between housing prices in different cities of Ukraine and the results may be interesting for regional and urban planning since implicit value of school quality and cost of crime is estimated, real estate experts and investors.

Chapter 2

LITERATURE REVIEW

This section starts from the discussion of spatial aspects of housing market, then reviews different estimation techniques and it concludes by a comparing of the results that these different models give when applied to real estate markets in some countries (Canada, Japan, the USA, France)

Traditionally, hedonic price analysis is applied to determine how certain characteristics of houses affect housing prices. The theoretical model for hedonic price analysis was developed by Rosen (1974). The model deals with heterogeneous good H (in this thesis it is a house, which is a heterogeneous good). Let's denote by P(H) an equilibrium price of a house $H = (H_1, H_2, \ldots, H_n)$, where H_i is a characteristic of a house H (for instance, total area or number of rooms etc.).

An individual maximizes one's utility subject to budget constraint:

$$U = \max [U(H,G)]$$

$$\text{s. t. } I = P(H) + GP(G)$$

I – income, G – other goods and P (G) – price of good G

The price paid for any characteristic of a good shows the utility an individual gets consuming a good with that characteristic.

Dubin (1988) outlines three groups of variables used in hedonic price models. The first one is structural variables, such as the size of a house, number of rooms etc. The second group is location variables, for instance, distance to the Central

Business District or to disadvantaged districts. The third category is neighborhood variables. This work includes structural and location variables explicitly in the model, while neighborhood variables will have affect on housing price through the prices of neighbor houses, since it is not always easy to measure these variables. Empirical findings of Long et al. (2007) suggest that while the traditional hedonic model is negatively affected by omission of neighborhood variables, spatial hedonic models still perform well.

Until the nineties traditional hedonic modeling was the prevalent method in studying the determinants of housing prices and mainly a linear specification form was used.

The problem with the traditional method is the assumption that housing characteristics have the same effect on housing prices in different areas, which contradicts the empirical findings. For instance, Bitter et al. (2007) using spatial expansion method and geographically weighted regression (GWR) showed that in Tucson housing market (Arizona) housing characteristics are valued differently. Yu (2004) using GWR found that "such house attributes as floor size, number of bathrooms, air conditioners and fire-places are more valued in rich districts of Milwaukee city (Michigan)".

The reason of varying price may be due to preferences of inhabitants for access or amenity values (Karlsson (2007)). If inhabitants change their preferences about housing location or housing characteristics, real estate market may face disequilibrium – the demand for certain types of houses will be unsatisfied; therefore, a price for highly demanded houses would rise, ensuing spatial heterogeneity in the market (Bitter et al. (2006)). Housing prices tend to decline with each mile as one moves further from the central business district (CBD) in Chicago (McMillen (2002)) or from the capital area in Iceland (Karlsson (2007)), however in the latter case 85 kilometers from the CBD housing prices start to increase. A possible explanation is the migration to suburb area, where a combination of access and amenity values is available.

One of the most widely used methods, which allow housing characteristics' "prices" to vary across space is geographically weighted regression (GWR) introduced by Brunsdon et al. (1998). This method allows to estimate the parameters for each geographical location *i*, "without specifying a functional form of spatial variation". The main difference between GWR and OLS is that in the former case parameters' estimates take into account data on neighbor houses. Geographically Weighted Regression assigns different weights to the prices of neighboring houses on the basis of their distance relatively to the point *i*.

Very similar to GWR is Moving Windows Regression. The main idea of Moving Windows Regression is to construct the window around point *i* and include in the regression the neighbors, which are located within this window (Long et al. (2007)).

Since the empirical findings indicate that housing characteristics are valued differently as one move across space, the logical approach would be to divide the housing market into relatively homogeneous submarkets on the basis of cluster analysis and calculate separately the contribution of housing characteristics to housing price. Then the results could be compared across the submarkets as was suggested by Case (2003). For the same purpose Tsutsumi et al. (2005) for Tokyo market and Mavrodiy (2005) for Kyiv market used dummy variables for every district to see whether geographical location affects housing prices, the major shortcoming is that this method does not show how valuations of different housing characteristics vary across space.

Typically the methods designed to cope with spatial heterogeneity (for instance, GWR, MWK) are computationally intensive and require not only special software, but also an additional set of variables such as the longitude and the latitude for each house listed for sale. Since such data is not available for Ukraine, I am going to test for the presence of spatial heterogeneity with the help of dummies included for each city.

However, spatial heterogeneity is not the only drawback of traditional hedonic price analysis. One of the reasons why the traditional model might not perform well is the assumption of no spatial correlation. In other words, it was assumed that two neighbor objects do not affect each other. As Dubin (1988) points out before nineties the researches either assumed the absence of spatial correlation or admitted the presence of autocorrelation and suggested it as a new direction for research. Prices of neighboring houses may be correlated because they have practically identical surrounding environment or because house buyers and sellers take into consideration neighbor housing prices when deciding upon the price they are willing to suggest or accept. What is more very often closely located houses are built at the same time and in the same manner, even if this is not a case, according to the conformity principle employed by appraisers if a luxury house is built in not prestigious area its price is below the price of a similar house in more prestige district (Besner (2002)).

Nowadays several approaches were developed to incorporate spatial dependency in traditional hedonic price analysis. Spatial hedonic modeling can be divided into two parts – geostatistics approach and spatial econometrics approach (Tsutsumi et al. (2005)).

In order to choose between different methods, I will briefly discuss the most widely used models. First, I will focus attention on theoretical aspects, which constitute the core of the models, and then the models will be compared according to their explanatory and predictive power.

Geostatistics approach deals with kriging models. A well-known kriging model was developed by Dubin (1988). He suggested computing a covariance matrix of error terms and using it for obtaining more accurate estimates. Maximum likelihood estimator was proposed to calculate simultaneously regression coefficients and covariance matrix elements employing the assumption that correlation between two points depends negatively on distance between them (negative exponential function) and that the error term is stationary. The last assumption is very strong because it implies that variance and mean do not change with the location.

Kriging is a very useful tool, because it makes estimates more precise through two channels: it incorporates spatial correlation in error terms into parameters estimates and it improves the prediction power by adding the predicted residual (calculated as a weighted average of estimated residuals) (Long et al. (2007)).

Kriging got further development and some modification by Haas (1995), his model was named Moving Windows Kriging (MWK). Unlike traditional kriging, MWK uses only the nearest neighbors for calculation of covariance matrix, so there is no need for the spatial stationarity assumption. The major advantage of MWK is its possibility to take into consideration both features of real estate market – spatial heterogeneity and spatial dependency (Long et al. (2007)

The Local Regression Model developed by Clapp (2002) emphasizes the time dimension of estimation by producing the prediction for housing price index controlled for quality at different points in time. For this purpose variables for latitude, longitude and time are included in the model.

Kriging models and the Local Regression Model can be used for the analysis of one city or one city district housing market because they take into account spatial dependence between the nearest neighbors. However, this is not the point of interest for my thesis, because I am testing the presence of spillover effect between different cities. In other words I am looking if there is spatial dependence between housing prices in different cities and if spatial correlation is more persistent in housing prices if two oblasts border each other or it is more important to be closely located not in terms of the geography, but in economic or demographic sense. I include structural characteristics of apartments in order to control for quality and dummies for all cities are expected to capture city specific effect. For this purpose, spatial econometrics techniques are useful.

Two the most popular models in spatial econometrics are a spatial lag model and a spatial error model (Baumont (1999)). Both models take into account spatial correlation, but a spatial lag model includes a spatial lag of endogenous variable because the price of an apartment is affected by the prices of neighboring apartments, while in a spatial error model spatial correlation is due to misspecification problems such as omitted variable or wrong functional form. An example of a spatial lag model would be Spatial Autoregressive Variable with Similarity components (SARS) developed by Besner (2002). The author models spatial correlation resulting from an individual's valuation of a certain house, which is affected by the prices of neighboring houses with the help of a linear autoregressive hedonic model, where the price of a more similar neighboring house has higher influence according to the weighting matrix, though closeness is superior to similarity characteristics.

The third approach incorporates both models – spatial autoregressive model and spatial error model and is called spatially autoregressive model with spatially correlated error terms (Kelejian et al. (1997)).

Practically all described models use a weighting matrix. Baumont (1999) discerns three types of weights – based on contiguity, distance and nearest neighbors. Case et al. (1993) adds one more type of weighting matrix – it can be based on socio-economic or demographic characteristics. Many authors (Long et al. (2007), Baumont (1999), Brunsdon et al. (1998), Bitter et al. (2006)) use nearest neighbor weighting utilizing different modifications of a Gaussian distance - decay function. This weighting scheme is applicable to calculate the model within the city, though it is computationally intensive and generally requires special software such as GWR 3.0, therefore, it is beyond the scope of my thesis.

As a result, other alternative ways to weighting will be applied. I will implement a weighting scheme based on contiguity, an inverse measure of the distance and the demographic characteristics of the city. To measure influence of one region (district) on another a simple weighting scheme is adopted (for example, Brady (2007)): $w_{ij} = 1$ if regions or states border each other and 0 – otherwise. The logic behind two other weighting schemes is also quite simple – if two cities are separated by larger distance (geographic distance or their population is very different in size), they are expected to affect each other less.

To be able to understand which econometric model is more appropriate for spatial hedonic modeling, it is necessary to discuss the empirical results and compare the models according to their out-of-sample prediction power and explanatory power. There is no consensus about the performance of the hedonic spatial models. Mainly the researchers agree that the traditional hedonic model gives worse results than models that take into account spatial correlation, but some results contradict to this statement.

Empirical evidence on performance of models developed by Case (division of the market into relatively homogeneous submarkets), Clapp (kriging version of local regression model) and Dubin (localized kriging model) can be compared on the basis of the results of their competition, which are described in Case et al. (2003). The criterion of effectiveness was the accuracy of out of sample prediction. The major result was that Case's and Dubin's models gave better results than OLS and Clapp's model.

Similar results regarding kriging model were obtained by Long et al. (2007). According to their findings kriging is the most robust model, GWR has high predictive power, MWR also gives good results and can be easier implemented comparatively to GWR, but MWK does the worst in terms of prediction. Bitter et al. (2006) confirms that GWR and the spatial expansion perform better than traditional hedonic model, though spatial expansion model is inferior to GWR if to compare accuracy of prediction and explanatory power.

Using the data set for Montreal housing market Besner (2002) found that adding of autoregressive parameter (SAR and SARS models) improves significantly the prediction power of the model comparatively with traditional model, however, SARS outperforms SAR. A weakness of these empirical results is low volatility of data as it is admitted by the author. Highly volatile Ukrainian data can be used to check the performance of SAR.

On the other hand, on the basis of prediction power Gao et al. (2002) came to the conclusion that neither GWR nor spatial dependency model performed better than traditional linear hedonic model for the housing market of Tokyo. Such outcome is due to small degree of spatial dependency in the data set or model misspecification.

This thesis uses four models to estimate how housing characteristics influence housing prices – traditional hedonic price model, spatial lag model (to see if spatial correlation is an attribute of housing prices), spatial error model and spatially autoregressive model with spatially correlated error terms. However, these models will be modified in order to fit not only within city analysis but also between cities analysis.

There are only few works about Ukrainian real estate market on micro level. The most relevant to the current study are works by Chekmezova (2007), Sioma (2006) and Mavrodiy (2005), but their works concentrate on Kyiv housing market, while this work expands the analysis for 8 cities of Ukraine. Chekmezova (2007) estimated a traditional hedonic model for housing market in Kyiv augmenting it with the level of pollution in order to determine marginal price for clean air, while Mavrodiy (2005) included in the regression dummies for administrative districts of Kyiv and found them to be significant. This thesis uses traditional hedonic model mainly as a benchmark and focuses on the spatial aspects, but it differs from Sioma (2006), since he used commuting time as a major factor that determined rental price differentials, while this work assumes that commuting time is one of the possible factors that explain housing heterogeneity, but others such as school quality, crime rate and average income may also be important determinants.

Chapter 3

DATA DESCRIPTION

The estimation is done for two aggregation levels – the relationship between housing prices across eight Ukrainian cities and ten districts of Kyiv is explored.

I use the data on housing prices and individual housing characteristics from 8 cities of Ukraine: Kyiv, Lviv, Rivne, Odesa, Kharkiv, Zhytomyr, Cherkasy and Dnipropetrovsk. The choice of the cities was determined by the factors described below.

These are the largest Ukrainian cities located in different parts of the country and have different socio - and demographic characteristics (the geographical location of the cities can be viewed below in Figure 3.1.). The similarity between them is that they are all administrative centers of the region (oblast). Finally, these are the only cities for which the data on apartment prices and individual housing characteristics is available.

Table 3.1: Distance between the cities (kilometers)

	Cherkasy	Dnipropetrovsk	Kharkiv	Kyiv	Lviv	Odesa	Rivne	Zhytomyr
Cherkasy		326	415	201	717	453	536	352
Dnipropetrovsk	326		222	479	948	463	814	630
Kharkiv	415	222		487	1042	685	823	638
Kyiv	201	479	487		544	480	324	140
Lviv	717	948	1042	544		793	215	407
Odesa	453	463	685	480	793		742	555
Rivne	536	814	823	324	215	742		187
Zhytomyr	352	630	638	140	407	555	187	

It is assumed that remoteness of the cities affects the degree of interrelation between housing prices – the greater is the distance (geographic, economic, demographic), the weaker is the dependence between housing prices and consequently a lower weight is given.

Figure 3.1: Map of Ukraine

As it was mentioned in the previous section, there are three groups of variables: structural (individual), location and neighbor. The individual characteristics in my data set are represented by variables listed in the Table 3.2.

These variables represent typical information listed for apartments on sale. Though I use seller price for the analysis, it is very close to the actual price of transaction, since on average the sellers are not willing to decrease the price by more than 3-5 %, some realtor agencies indicate that this number is even lower – about 1-1.5 %[3], which can be considered as a negligible amount.

[3] http://bin.com.ua/templates/analitic_article.shtml?id=77075

Table 3.2: Structural characteristics of an apartment

Variable	Type/ units of measurement	Description
Price	Log transformation of asking price, which was initially in USD	Asking price for one-room apartments
Total area	Square meters	Total area of an apartment listed for sale
Squared area	Squared total area	Squared area of an apartment listed for sale
Material	Dummy	1 – if the house is made of bricks 0 - otherwise
Floor	Dummy	1 – if an apartment is located on the first/last floor 0 - otherwise
Repair	Dummy	1 – if recently repair was made 0 - otherwise

The total area is expected to have a positive sign because typically individuals value larger apartments, but the squared area is expected to have a negative sign, since according to the theory each additional square meter brings less utility to an individual.

As is often mentioned by real estate experts typically houses made of bricks have better conditions than those made of concrete panel blocks, therefore the variable material is expected to be positively correlated with the price.

The variable floor is included to distinguish between the first/last floor and any other floor. Typically first/last floors are considered to be the worst due to higher probability of burglary, need to place grates on the windows, noise if an apartment is located on the first floor and often problems with the roof if an apartment is located on the last floor. Repair stands for improved interior conditions inside the apartment and it should lead to higher price.

Individual characteristics allow control for the quality of an apartment and their choice was determined by previous research (more detailed description is in the section "Literature review").

Location variables are represented by dummies for each city except Kyiv, because Kyiv is used as a reference group. They allow capture city specific effect.

Neighbor variables (available for the districts of Kyiv city only) [4] are presented by:

a) ratio of the number of students in district schools to the total number of children aged 6-18 in the district – this variable is a proxy of school quality in the district (higher ratio implies higher school quality);

b) ratio of total number of workplaces in the district to the population of the district – very low ratio will indicate that a district is a so-called "bedroom community", while very large ration means that a district is either a business center or industrial district;

c) changes in population due to migration per 1000 inhabitants;

d) ratio of the number of crimes committed in the district to the number of inhabitants in the district.

The data is available in the form of Multiple Listing Service (MLS) from newspaper Aviso, which is issued two times a week. However, if a house is listed for sale, it does not necessarily mean it was sold – it may be either withdrawn from the sale or not sold due to the absence of demand. But this is the only available source of information on Ukrainian housing prices on a regular basis;

[4] www.statyst.kyiv-city.gov.ua

therefore, we will assume that though an apartment was withdrawn from the sale, it reflected the market price.

A selection bias problem may arise (Baumont (1999)) because houses listed in the newspaper may not be representative houses for the district (city). However, since our data set is quite large, we expect to overcome this problem. Subjectivity of data is also a feature of Multiple Listing Service (Sioma (2006)) – sellers do not necessary include all relevant or true information (for instance, the information concerning the condition of an apartment). The information is given in a free form; therefore, some sellers include the information about the presence of telephone and balcony, while others – do not. To handle this issue, observations with missing values were excluded.

The sample contains data on housing prices and major housing characteristics for October 2007 (randomly chosen month).

We first start from the description of the housing prices for eight Ukrainian cities in table 3.3 and table 3.4.

Table 3.3: Descriptive statistics of prices and total area of apartments in Ukrainian cities (one-room apartment)

City	No of observations	Apartment price (in USD)				Total area (sq. meters)			
		Min	Max	Mean	St. dev	Min	Max	Mean	St. dev
Kyiv	2819	48000	850000	116899	40152.47	18	97.4	36.0133	8.0348
Lviv	436	20000	159300	57735.7	15358.6	15	195	35.6206	12.0437
Rivne	521	20000	75000	38267.2	7745.43	12	53	32.4203	5.46487
Zhytomyr	426	26000	85000	42168.3	5906.65	17	59	33.8915	4.48563
Cherkasy	610	15500	102000	36770.3	7953.08	18	57	33.3469	4.0952
Kharkiv	483	17000	85000	43008.3	10861.8	21	79	34.7737	6.64379
Dnipropetrovsk	522	21000	150000	49262.1	14671.9	19	66	34.3341	5.7966
Odesa	590	26000	260000	63259.4	23601.6	17	112	36.5349	9.5637

To visualize the descriptive statistics on the dependent variable, we included figure 3.2. It emphasizes the fact that Kyiv has much higher prices than other cities. From table 3.4 it can be seen that apartments have different structural characteristics, but these difference could explain difference between in housing prices between any of seven cities except Kyiv. Eyeball test allows us to make a statement that there is spatial heterogeneity and apartments are valued higher in Kyiv, but we cannot make any preliminary guess about spatial heterogeneity in housing market if seven other cities are considered.

It is quite difficult, however, to say a priori whether there is spatial dependence in the housing prices across cities.

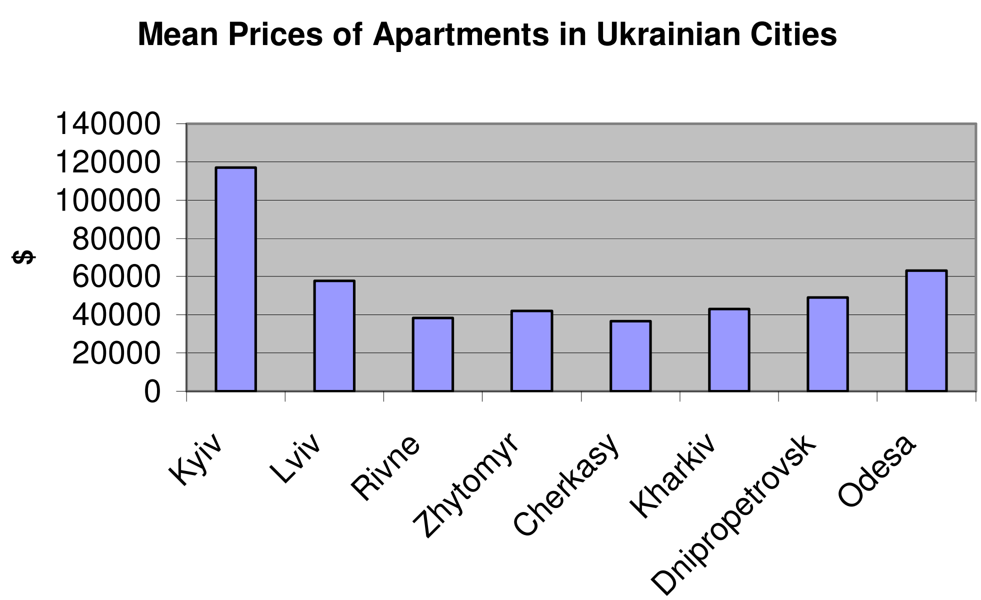

Figure 3.2: Mean Prices in Ukrainian Cities, October 2007

Table 3.4: Descriptive statistics of one-room apartments in Kyiv, Lviv, Kharkiv, Dnipropetrovsk, Rivne, Zhytomyr, Cherkasy, Odesa

City	Positive outcome, %		
	Type of material for the house is brick	An apartment is located on the lowest/highest floor	An apartment with repair
Kyiv	42	23	30
Lviv	78	50	30
Rivne	59	36	29
Zhytomyr	67	27	22
Cherkasy	48	32	17
Kharkiv	65	36	25
Dnipropetrovsk	21	27	27
Odesa	19	30	33

Since Kyiv has the largest number of houses for sale, the largest area and the highest population, it is pertinent to consider separately a model for districts of Kyiv.

Table 3.5: Descriptive statistics of price and total area of one - room apartments in the districts of Kyiv

District	No of observations	Apartment price (in USD)				Total area (sq. meters)			
		Min	Max	Mean	St. dev.	Min	Max	Mean	St. dev.
Golosiyivskiy	260	50000	265000	122169	39662.7	21	66	38.5604	10.2429
Darnytskiy	210	59700	200000	122183	29440.4	20.7	74	44.063	9.0299
Desnyanskiy	323	63000	255000	92096	25271.2	21	65.1	34.2317	8.53971
Dniprovskiy	299	52000	350000	96927.4	27267.5	19	66.7	34.0152	6.38264
Obolonskiy	343	57000	280000	110688	40213.7	19	84	35.3367	7.9851
Podilskiy	169	62000	400000	106808	54081.5	20	76	33.2864	8.58347
Pecherskiy	186	79000	499000	167118	72812	19	80	35.9511	8.83794
Shevchenkivskiy	360	63000	850000	134137	86410.1	20	97.4	34.5953	9.65959
Solomyanskiy	288	52000	210000	102581	30290.8	19	61.6	34.1104	8.44309
Svyatoshynskiy	387	48000	718000	93548.8	38510.9	18	64.7	34.3424	7.35783

As it can be inferred from the table 3.5, on average Pecherskiy district contains the most expensive apartments in Kyiv. The lower bound for the area of an apartment is quite similar in all the districts – about 20 sq. meters. On average the size of an apartment does not differ significantly across districts.

Table 3.6: Descriptive statistics for structural characteristics of 1-room apartments in the districts of Kyiv

District of Kyiv	Positive outcome, %					
	Type of material for the house is brick	An apartment is located on the lowest/highest floor	Presence of a balcony/loggia	Presence of the telephone	An apartment with repair	An apartment has average conditions for living
Golosiyivskiy	86	22	98	29	27	59
Darnytskiy	51	17	99	36	48	43
Desnyanskiy	74	26	97	32	30	66
Dniprovskiy	61	23	98	38	27	67
Obolonskiy	41	24	97	39	29	64
Podilskiy	81	26	98	39	28	67
Pecherskiy	93	23	98	26	34	54
Shevchenkivskiy	91	31	92	30	37	58
Solomyanskiy	82	27	92	34	33	60
Svyatoshynskiy	72	19	94	34	39	51

The difference in such structural characteristics of houses as the floor (the highest/lowest), presence of balcony/loggia, presence of the telephone, condition of apartment (recent repair or average condition for living) is within 10%. In other words, apartments for sale are quite similar in different districts of Kyiv. Though structural characteristics affect the price of each individual apartment, they do not lead to different prices in different districts of the same city, unless houses/apartments differ across districts.

Neighbor variables (demographic and socio-economic) are expected to cause difference in prices for the same type of apartments, which are located in different places and they are summarized in table 3.7.

Table 3.7: Demographic and socio-economic situation in the districts of Kyiv

Districts of Kyiv	Population (01.07.07)	Number of crimes committed	Change in the population due to migration per 1000 of inhabitants	Children at the age 6-18 in the district	Number of students at schools	Ratio of workplaces in the district to the total population in the district, %
Golosiyivskiy	220735	218637	-0.1	17397	19282	52.26
Darnytskiy	301634	295135	5.1	32788	29245	17.19
Desnyanskiy	348683	345793	3.6	33079	30581	11.90
Dniprovskiy	339519	337790	5.9	28936	32388	19.19
Obolonskiy	311309	308435	3.0	30871	27155	24.55
Pecherskiy	134628	129905	-0.3	7433	14529	150.48
Podilskiy	186136	182726	4.7	14733	17724	41.76
Svyatoshinskiy	326787	320867	4.4	29804	29737	26.43
Solomyanskiy	320314	318268	7.7	21895	24487	42.05
Shevchenkivskiy	231994	222876	7.1	18094	23395	106.36

Most of the districts attract newcomers. Only two districts – Golosiyivskiy and Pecherskiy – exhibited decrease in population, which is very small in size and can be considered as negligible. Decrease in population may be caused by either more expensive apartments or non good conditions in the district. Therefore, its impact on housing prices is ambiguous.

If a district has more students at schools than total number of children aged 6-18, it testifies high quality of schools and they are expected to be one of the contributors to high housing prices in the district.

Ratio of the number of workplaces in the district to the total number of population shows to some degree whether a district is a designed as a bedroom community or it is business center or industrial district (if the and we cannot say exactly which sign we expect due to the difference in preferences of the individuals.

People also value safety and are willing to live in the districts with as little crimes committed as possible.

Chapter 4

METHODOLOGY

A necessary attribute of each house is its location, which partially determines its value. Spatial econometrics is a field that allows incorporate spatial features (spatial dependency and spatial heterogeneity) into regression models.

Usual OLS is not appropriate, since it leads to inefficient results in some cases or biased and inconsistent results in the other cases. (Anselin (1988)). However, traditionally, OLS is used as a benchmark and residuals from OLS regression are used in the tests for spatial heterogeneity and spatial dependence.

Pace et al. (1998) indicate two ways how spatial dependency is traditionally handled. The first approach is to include an additional set of variables in traditional hedonic pricing model (for instance, besides individual housing characteristics, location and neighbor characteristics are included); the goal is to make error terms patternless across space. The second way is to model the dependence between residuals.

We start from estimating traditional hedonic model including only individual characteristics of an apartment and dummy for each city. Kyiv is a reference group. We assume no spatial dependency or heterogeneity is present in real estate market.

The first model is *traditional hedonic model* :

$$\ln(HP) = \beta_1 + \beta_2 I + \beta_3 L + \beta_4 N + \varepsilon \tag{4.1}$$

$$\varepsilon \sim N(0, \sigma^2 I)$$

ln(HP) – logarithm of housing price

I – vector of individual characteristics

L – vector of location characteristics

N – vector of neighbor characteristics

The choice of semi log functional form is based on the work of Chekmezova (2007), where empirical findings for Kyiv housing markets showed that semi log model performed better than linear form.

Further the results are explored for the presence of spatial heterogeneity. This is done by examining the dummies for cities, which are included into the model (reference group is Kyiv). Significant dummies indicate difference in valuation of apartments across the cities.

The next step is to test for the presence of spatial dependence in housing prices for different cities. Spatial dependence may be represented in the form of spatial lag in the model, spatial error correlation or the above two combined.

A spatial lag means that the price of an apartment is directly affected by prices in the neighbor cities, possible because an individual observes these prices and anticipates housing price in his/her native town to follow the same path. Another reason is that too high prices in some city (for instance Kyiv) may force individuals to move to similar cities in terms of socio-economic characteristics or those, which are located closer in terms of geography, consequently, the demand for apartments in those cities will increase and rise of housing prices will be observed. This is the so-called spillover effect in real estate market. Omitting spatial lag will lead to biased and inconsistent results.

Spatial error correlation is less severe problem than omitted spatial lag, since the consequences are unbiased but inefficient estimates and it can be eliminated if additional set of neighbor variables is included into the model. The rationale behind spatial error correlation is some variable, which affects housing prices in the cities is omitted and this leads to correlation of error terms.

One model can contain both a spatial lag and spatial error correlation present.

Spatial models require exogenously given weighting matrices. With the help of weighting matrices we express our a priori beliefs about housing prices in which cities are stronger correlated by assigning higher weights for those cities. The matrices can be of three types (Case et al. (1993)):

- Based on distance – the inverse distance serves as a weight, thus, weight for more distant cities is lower than for geographic neighbors;

- Based on contiguity – if two regions (we use cities - regional centers) border each other weight is 1 and it is 0 otherwise;

- Based on socio-economic or demographic characteristics of the oblasts – inverse of difference in population is used as a weight; cities with similar number of inhabitants are considered to be neighbors and their housing prices are expected to be interrelated.

Traditionally, weights are normalized and their row sum is equal to 1. Diagonal entries of a matrix are equal to 0.

In order to determine which form of spatial dependence is present, LM test for the presence of spatial lag and spatial error is a useful technique.

Lagrange Multiplier test for the presence of spatial lag:

$$LM_{lag} = [Ne'Wy / e'e]^2 [N(WX\beta)' M(WX\beta) / e'e + tr(W'W + W^2)]^{-1} \qquad (4.2)$$

N – number of observations, W – weighting matrix, e - error terms after OLS regression, y – vector of dependent variable, X – matrix of independent variables, β - coefficients from OLS regression, tr – means trace, $M = I - X(X'X)^{-1}X'$

Lagrange Multiplier test for the presence of spatial error correlation:

$$LM_{error} = [Ne'We / e'e]^2 [tr(W'W + W^2)]^{-1} \qquad (4.3)$$

In both cases null hypothesis is the absence of particular from of spatial dependence. In other words, under the null OLS specification is appropriate. The statistics on both tests is distributed as chi-square with 1 degree of freedom.

The above tests have some drawbacks. They are not robust to the alternative form of spatial dependence, that is, though in fact there is only spatial lag present, both LM test for spatial lag and LM test for spatial error correlation will reject the null and or if we have only spatial error correlation LM test for spatial error correlation rejects the null and what is more LM test for spatial lag also rejects the null and instead of spatial error model (SEM), spatial autoregressive model with spatially correlated error terms will be estimated. What is more, while LM test for spatial lag is robust to heteroskedasticity and non-normally distributed error terms, LM test for spatial error correlation underrejects the null if error terms are not normally distributed and overrejects in the presence of heteroskedasticity.

As a result, robust LM tests for spatial lag and spatial error correlation were developed. They are robust to alternative form of spatial dependence; however, they tend to underreject the null hypothesis.

Robust Lagrange Multiplier test for the presence of spatial lag:

$$LM_{lag}^{robust} = \frac{(e'Wy/s^2 - e'We/s^2)^2}{(NJ_{\rho\beta}) - tr(W'W + W^2)} \tag{4.4}$$

$$s^2 = \frac{e'e}{N}, \ (NJ_{\rho\beta})^{-1} = [tr(W'W + W^2) + (WX\beta)'M(WX\beta)/s^2]^{-1}$$

All other notation is the same as for the non-robust Lagrange Multiplier test for the presence of spatial lag.

Robust Lagrange Multiplier test for the presence of spatial error correlation:

$$LM_{error}^{robust} = \frac{[(e'We/s^2 - tr(W'W + W^2)(NJ_{\rho\beta})^{-1}(e'Wy/s^2)]^2}{tr(W'W + W^2) - [tr(W'W + W^2)]^2(NJ_{\rho\beta})^{-1}} \tag{4.5}$$

Notation remains the same as in the above LM tests.

Robust LM tests also have chi-square distribution with 1 degree of freedom.

Common practice is to calculate all four tests, compare the results and conclude which type of spatial model is appropriate.

If null hypothesis is rejected on the basis of the results of *LM test for spatial error correlation*, the second model to be estimated is *spatial error model* (SEM):

$$\text{Ln(HP)} = \beta_1 + \beta_2 \text{I} + \beta_3 \text{L} + \beta_4 \text{N} + \varepsilon \tag{4.6}$$

$\varepsilon = \lambda W\varepsilon + u$

λ – parameter, which shows how strong is spatial correlation

W – weighting matrix

Unlike OLS in the spatial error model the residuals are spatially correlated and it has to be taken into account when choosing the estimation technique. Spatial correlation between error terms is viewed as a result of omitted variables (either location or neighboring variables, which are not always possible to measure).

Off-diagonal elements of variance-covariance matrix are non-zero in SEM. According to Anselin (2006) solving for error terms yields: $\varepsilon = (I - \lambda W)^{-1} u$

SEM model is equivalent to:

$$\ln(HP) = \lambda W \ln(HP) + X\beta - \lambda WX\beta + \varepsilon \qquad (4.7)$$

and variance-covariance matrix is:

$$E(\varepsilon\varepsilon') = \sigma^2 (I - \lambda W)^{-1} (I - \lambda W')^{-1} \quad because\ E(uu') = \sigma^2 I \qquad (4.8)$$

If the number of neighbors for each observation differs, heteroskedasticity will be present in the model. OLS can be applied to estimate SEM if error terms are adjusted for spatial correlation.

Another method to estimate the model with spatially correlated error terms is GMM. The main advantages of this method are the absence of the normality assumption for the distribution of error terms and the ease of implementation with standard software such as STATA.

If the null hypothesis is rejected on the basis of the results of *LM test for spatial lag*, the third model to be estimated is *spatially autoregressive model* (SAR):

$$\ln(HP) = \alpha_1 + \varrho W\ln(HP) + \alpha_2 I + \alpha_3 L + \alpha_4 N + \varepsilon \tag{4.9}$$

ϱ – parameter, which measures spatial dependence of housing prices;

$W\ln(HP)$ – spatial lag, which is in fact weighted average of logarithmically transformed prices in neighbor cities.

When we estimate ln(HP) for object i, we will include information about neighbor object j, at the same time doing computation for object j, we will include information about neighbor object i. This has to be taken into account, when choosing estimation procedure.

Following Anselin (2006) the reduced form, which contains no spatially lagged term on the right side is presented as:

$$\ln(HP) = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\varepsilon \tag{4.10}$$

It can be inferred from the above expression that housing price depends not only on individual characteristics of certain house, but also on characteristics of neighboring houses.

After expansion of the inverse term and taking expectations we would get:

$$E(\ln(HP) \mid X) = X\beta + \rho WX\beta + \rho^2 W^2 X\beta + \ldots \tag{4.11}$$

In this situation ϱ has to be computed simultaneously with other parameters. Two possible estimation techniques are Maximum Likelihood and Method of Moments.

The major assumption for ML estimation is normality of error terms.

Log-likelihood function is the following:
$$L = -(N/2)(\ln 2\pi) - (1/2)\ln\sigma + \ln|I - \rho W| - (1/2)(\ln HP - \rho W \ln HP - X\beta)'\sigma^{-1} \times$$
$$\times(\ln HP - \rho W \ln HP - X\beta) \tag{4.12}$$

Method of moments does not require normality of error terms; they only have to be i.i.d. distributed. Since spatial lag of dependent variable Wln(HP) is endogenous, it has to be instrumented. Spatial lags of independent variables WX can serve as instruments in this situation.

If null hypothesis is rejected on the basis of the results of *LM test for spatial error correlation* as well as *LM test for spatial lag*, the fourth model to be estimated is <u>*spatially autoregressive model with spatially correlated error terms*</u> (I follow Kelejian et al. (1997) in formulation and estimation procedure for the model):

$$Ln(HP)_n = X_n\beta + \varrho W_n(HP)_n + u_n \qquad |\varrho|<1 \tag{4.13}$$

$$u_n = \lambda M_n u_n + \varepsilon_n \qquad |\lambda|<1$$

W and M are weighting matrices, but the most common practice is to take the same weighting matrix W=M.

Three step procedure is appropriate for estimation of the model. Detailed description of each step and necessary assumptions can be found in Appendix A.

In the next section of the thesis, I will present the results of the specification tests and compare OLS with spatial models.

C a p t e r 5

ESTIMATION RESULTS

We will start from the estimation of OLS, which will be used as a benchmark to compare with spatial models. We assume that there is neither spatial dependence nor spatial heterogeneity, when we estimate this model.

Next presence of spatial heterogeneity will be tested.

H_0: no spatial heterogeneity

H_1: average prices of apartments differ across cities (different intercept for different cities) and/or valuation of characteristics differ (different slopes)

In the model Kyiv was used as a reference and dummies and interactions for other 7 cities were included.

Table 5.1: OLS estimation

floor	-0.030***	-0.028
	[0.007]	[0.021]
area	0.031***	0.053***
	[0.001]	[0.007]
repair	0.076***	0.130***
	[0.007]	[0.019]
material	0.057***	0.039**
	[0.007]	[0.017]
dumcherk	-0.992***	-1.014***
	[0.012]	[0.281]
dumdnipr	-0.738***	-0.192
	[0.012]	[0.195]
dumkharkiv	-0.864***	-0.097

Table 5.1: Cont.

	[0.012]	[0.178]
dumlviv	-0.613***	-0.108
	[0.013]	[0.137]
dumodesa	-0.536***	-0.078
	[0.012]	[0.147]
dumrivne	-0.953***	-0.636***
	[0.012]	[0.199]
dumzhytomyr	-0.876***	-0.441*
	[0.012]	[0.259]
areasq	-0.000***	-0.000***
	(0.000)	[0.000]
cherfloor		-0.007
		[0.028]
dniprfloor		0.078***
		[0.029]
kharkivfloor		-0.135***
		[0.028]
lvivfloor		0.033
		[0.027]
odesafloor		0.007
		[0.029]
rivnefloor		0.018
		[0.028]
zhytomyrfloor		0.013
		[0.028]
cherarea		0
		[0.016]
dniprarea		-0.031***
		[0.011]
kharkivarea		-0.035***
		[0.009]
lvivarea		-0.025***
		[0.007]
odesaarea		-0.026***
		[0.007]
rivnearea		-0.017
		[0.011]
zhytomyrarea		-0.017
		[0.015]
cherrepair		-0.071**
		[0.031]
dniprrepair		-0.067**
		[0.027]
kharkivrepair		-0.032

Table 5.1: cont

		[0.027]
lvivrepair		-0.041
		[0.028]
odesarepair		-0.04
		[0.026]
rivnerepair		-0.108***
		[0.027]
zhytomyrrepair		-0.072**
		[0.028]
chermater		-0.008
		[0.025]
dniprmater		0.054*
		[0.028]
kharkivmater		0.071***
		[0.026]
lvivmater		-0.001
		[0.028]
odesamater		0.039
		[0.028]
rivnemater		0.03
		[0.025]
zhytomyrmater		-0.005
		[0.025]
cherareasq		0
		[0.000]
dniprareasq		0.000***
		[0.000]
kharkivareasq		0.000***
		[0.000]
lvivareasq		0.000***
		[0.000]
odesaareasq		0.000***
		[0.000]
rivneareasq		0
		[0.000]
zhytomyrareasq		0
		[0.000]
Constant	10.550***	10.126***
	[0.025]	[0.124]
Observations	3408	3408
R-squared	0.80	0.81

Standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

The table above presents regression coefficients for two cases: when we allow all coefficients to vary across different cities (for this purpose interactions are included) and when only dummies for each city are included. In both cases Kyiv is used as a reference group.

Let's start form the discussion of the results for the model without interactions. The dummy floor has value 1 if an apartment is located on the first or the last floor (since psychologically people are not willing to choose apartments there), as it was expected the dummy has negative sign and it states that apartment on the first or the last floor will cost 3% cheaper that apartment of the same quality on the other floor. The other structural characteristics (area, a dummy for recent repair and a dummy for material, which takes value 1 if a house, where an apartment is located, is made of bricks) affect the price positively.

Comparing two models, we can see that coefficients differ significantly. For instance, area and repair are valued higher in Kyiv, while the opposite is true for the type of material for the house. Floor is not significant for the reference group in the model with interactions.

If we consider the model without interactions, the result is that all dummies are significant and have negative sign, which is consistent with our expectations that on average prices in Kyiv are higher than in other regions and we can confirm the hypothesis about spatial heterogeneity, in other words, not only structural characteristics matter, but also location. However, dummies for Zhytomyr and Kharkiv do not seem to be different.

When we take a look at the model with interactions most dummies for cities are insignificant. Only dummies for Cherkasy and Rivne remained significant (dummy for Zhytomyr is also significant but at 10 % level of significance). However, we can see that coefficients vary across cities. For instance, additional square meter is valued less in Kharkiv, Dnipropetrovsk, Lviv or Odesa than in reference group (Kyiv). The same statement is true for valuation of repair if we compare Cherkasy, Dnipropetrovsk, Rivne and Zhytomyr with Kyiv. Individuals in Kharkiv and Dnipropetrovsk are willing to pay higher price for apartments in the houses made of bricks if to compare with Kyiv. Therefore, not only average prices differ, but also different characteristics have different values if we move across the cities.

Thus, spatial heterogeneity is indeed the feature of housing market in Ukraine.

We proceed with spatial models to test the hypothesis about spatial dependence.

In order to perform spatial models I made the number of observations equal for each city (as a result we had randomly chosen 426 observations for each city) and we compared similar apartments in different cities (the criteria for similarity was total area).

The weighting matrices used are of three types:

- Based on distance
- Based on contiguity
- Based on demographic characteristics of the city.

For the first type inverse measure of distance between the cities is used as a weight. The rationale behind this measure is that cities located close to each other often have similar economic specialization and, therefore, are somewhat similar. For instance, cities in the eastern Ukraine (Kharkiv, Dnipropetrovsk) are traditionally industrial cities, and we expect housing prices in Dnipropetrovsk to have greater effect on housing prices in Kharkiv, than in Rivne or Lviv, which are located in the western Ukraine.

Initially weighting matrix was symmetric and each weight represented absolute measure of the inverse of distance between two cities as it is shown in table 5.2.

Table 5.2: Weighting matrix based on the inverse distance between the cities (non-normalized)

	Cherkasy	Dnipropetrovsk	Kharkiv	Kyiv	Lviv	Odesa	Rivne	Zhytomyr
Cherkasy	0	0.0031	0.0024	0.0050	0.0014	0.0022	0.0019	0.0028
Dnipropetrovsk	0.0031	0	0.0045	0.0021	0.0011	0.0022	0.0012	0.0016
Kharkiv	0.0024	0.0045	0	0.0021	0.0010	0.0015	0.0012	0.0016
Kyiv	0.0050	0.0021	0.0021	0	0.0018	0.0021	0.0031	0.0071
Lviv	0.0014	0.0011	0.0010	0.0018	0	0.0013	0.0047	0.0025
Odesa	0.0022	0.0022	0.0015	0.0021	0.0013	0	0.0013	0.0018
Rivne	0.0019	0.0012	0.0012	0.0031	0.0047	0.0013	0	0.0053
Zhytomyr	0.0028	0.0016	0.0016	0.0071	0.0025	0.0018	0.0053	0

Next this matrix is normalized: $w_{ij}^{norm} = \frac{w_{ij}}{\sum_i w_{ij}}$

w_{ij}^{norm} - element of normalized weighting matrix

w_{ij} - element of non-normalized weighting matrix

After normalization procedure weighting matrix is non-symmetric and its elements show relative distance between cities, taking into account how many near neighbors each city has. Let's say, if we measure the absolute inverse distance between Cherkasy and Dnipropetrovsk it is equal to 0.0031 (we can see it from table 2); if we look at table 3 with normalized weighting matrix it is noticeable that Cherkasy housing prices are assumed to have higher influence on Dnipropetrovsk housing prices than the relationship in other direction. This is because Cherkasy is located in the Central part of Ukraine and consequently cities located in Western Ukraine (Lviv, Rivne) and Central Ukraine (Zhytomyr, Kyiv) are relatively closer to Cherkasy than to Dnipropetrovsk (which is located in Eastern Ukraine and only cities in Eastern Ukraine represented by Kharkiv are geographically closer to Dnipropetrovsk than to Cherkasy), in other words Cherkasy has more near neighbors, therefore, each neighbor gets lower weight. Thus, standardized weights not only show distance between cities but also take into account how distant other neighbors are.

Table 5.3: Weighting matrix based on inverse distance between the cities (normalized)

	Cherkasy	Dnipropetrovsk	Kharkiv	Kyiv	Lviv	Odesa	Rivne	Zhytomyr
Cherkasy	0	0.1635	0.1284	0.2652	0.0743	0.1177	0.0994	0.1514
Dnipropetrovsk	0.1955	0	0.2871	0.1331	0.0672	0.1377	0.0783	0.1012
Kharkiv	0.1701	0.3179	0	0.1449	0.0677	0.1030	0.0857	0.1106
Kyiv	0.2138	0.0897	0.0873	0	0.0790	0.0895	0.1327	0.3070
Lviv	0.1024	0.0775	0.0705	0.1350	0	0.0926	0.4316	0.1804
Odesa	0.1792	0.1753	0.1185	0.1691	0.1023	0	0.1094	0.1462
Rivne	0.0995	0.0655	0.0648	0.1647	0.2482	0.7191	0	0.2853
Zhytomyr	0.1249	0.0698	0.0689	0.3140	0.1080	0.0792	0.2351	0

The normalization procedure was applied to other weighting matrices as well and the normalized matrices are presented in tables 5.4 and 5.5.

For the second type equal weight is given to the oblasts that share common border with the oblast of interest. The logic is similar to the inverse measure of distance; however, we assume housing prices in western Ukraine have no direct influence on the housing prices in eastern Ukraine. However, since only 8 cities are included in the regression analysis, some cities (Odesa) do not have neighbors in the dataset.

Table 5.4: Weighting matrix based on the contiguity (normalized)

	Cherkasy	Dnipropetrovsk	Kharkiv	Kyiv	Lviv	Odesa	Rivne	Zhytomyr
Cherkasy	0	0	0	1	0	0	0	0
Dnipropetrovsk	0	0	1	0	0	0	0	0
Kharkiv	0	0	1	0	0	0	0	0
Kyiv	0.5	0	0	0	0	0	0	0.5
Lviv	0	0	0	0	0	0	1	0
Odesa	0	0	0	0	0	0	0	0
Rivne	0	0	0	0	0.5	0	0	0.5
Zhytomyr	0	0	0	0.5	0	0	0.5	0

For the third type the inverse of difference in the city population serves as a weight. We assume that number of inhabitants in the city is a good indicator of presence of workplaces, wages, and educational opportunities due to migration. Therefore, if two cities have similar size of the population, we expect cities to be similar and housing prices in these cities to be correlated.

Table 5.5: Weighting matrix based on the inverse of difference in the number of inhabitants in the cities (normalized)

	Cherkasy	Dnipropetrovsk	Kharkiv	Kyiv	Lviv	Odesa	Rivne	Zhytomyr
Cherkasy	0	0.0115	0.0077	0.0039	0.0196	0.0123	0.1909	0.7540
Dnipropetrovsk	0.0434	0	0.0873	0.0223	0.1057	0.6576	0.0409	0.0428
Kharkiv	0.0885	0.2665	0	0.0912	0.1459	0.2353	0.0851	0.0876
Kyiv	0.1083	0.1639	0.2199	0	0.1354	0.1585	0.1062	0.1078
Lviv	0.1444	0.2071	0.0937	0.0360	0	0.2468	0.1310	0.1408
Odesa	0.0457	0.6460	0.0757	0.0211	0.1237	0	0.0429	0.0449
Rivne	0.3834	0.0218	0.0149	0.0077	0.0357	0.0233	0	0.5133
Zhytomyr	0.7089	0.0107	0.0072	0.0037	0.0179	0.0114	0.2402	0

Having estimated OLS regression and specified weighting matrix, several specification tests are performed to see if there is spatial correlation in error terms, need for spatial lag of dependent variable or both.

Null hypothesis: OLS is an appropriate model and no spatial dependence of any form is present in the model

Alternative hypothesis: spatial dependency (in the form of spatial lag or spatial error) is present

The results of robust LM tests for spatial lag and spatial error correlation (Appendix B) indicate the presence of both spatial error correlation and spatial lag in the model; therefore, spatially autoregressive model with spatially correlated error terms seems to be the best option. But for the sake of comparison spatial error model and spatially autoregressive model will be estimated as well.

As a result, three models were estimated: spatial error model (SEM), spatially autoregressive model (SAR) and spatially autoregressive model with spatially correlated error terms.

- *Spatial error model*:

In spatial error model the coefficients look quite similar and do not seem to differ as we change the type of weighting matrix. The model does not outperform OLS in terms of explanatory power. Moreover, according to the theory the coefficients from spatial error model do not differ much from the OLS coefficients even in the presence of spatial correlation. However, since spatial correlation is present, the standard errors differ from those obtained after OLS regression, because OLS estimates are inefficient.

Table 5.6: Estimation results for spatial error model (SEM)

lnprice	Weighting matrix based on the inverse of distance	Weighting matrix based on contiguity	Weighting matrix based on the demographic characteristics
floor	-0.030***	-0.030***	-0.029***
	[0.007]	[0.007]	[0.007]
area	0.031***	0.031***	0.031***
	[0.001]	[0.001]	[0.001]
repair	0.076***	0.076***	0.077***
	[0.007]	[0.007]	[0.007]
material	0.056***	0.057***	0.056***
	[0.007]	[0.007]	[0.007]
areasq	-0.000***	-0.000***	-0.000***
	[0.000]	[0.000]	[0.000]
dumcherkasy	-0.992***	-0.992***	-0.992***
	[0.012]	[0.012]	[0.012]
dumdnipropetrovsk	-0.738***	-0.738***	-0.738***
	[0.012]	[0.012]	[0.013]
dumkharkiv	-0.864***	-0.864***	-0.865***
	[0.012]	[0.012]	[0.013]
dumlviv	-0.613***	-0.613***	-0.613***
	[0.013]	[0.013]	[0.013]
dumodesa	-0.537***	-0.438***	-0.537***
	[0.013]	[0.012]	[0.013]
dumrivne	-0.953***	-0.953***	-0.953***
	[0.012]	[0.012]	[0.012]
dumzhytomyr	-0.876***	-0.876***	-0.876***
	[0.012]	[0.012]	[0.012]
Constant	10.775***	10.454***	11.286***
	[0.025]	[0.025]	[0.025]
Observations	3408	3408	3408
R-squared	0.8	0.81	0.81

Standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Spatially autoregressive model (SAR):

Table 5.7: Estimation results from spatially autoregressive model

Lnprice	Weighting matrix based on the inverse of distance	Weighting matrix based on contiguity	Weighting matrix based on demographic characteristics
wlnprice	**0.430*** **	**0.254*** **	**0.368*** **
	[0.046]	**[0.031]**	**[0.054]**
Floor	-0.030***	-0.031***	-0.031***
	[0.007]	[0.007]	[0.007]
Area	0.018***	0.025***	0.021***
	[0.002]	[0.001]	[0.002]
Repair	0.073***	0.070***	0.073***
	[0.007]	[0.007]	[0.007]
material	0.058***	0.057***	0.056***
	[0.007]	[0.007]	[0.007]
Areasq	-0.000***	-0.000***	-0.000***
	[0.000]	[0.000]	[0.000]
dumcherkasy	-1.132***	-1.249***	-0.968***
	[0.019]	[0.034]	[0.013]
dumdnipropetrovsk	-0.797***	-0.764***	-0.811***
	[0.014]	[0.013]	[0.017]
dumkharkiv	-0.938***	-0.920***	-0.913***
	[0.015]	[0.014]	[0.015]
dumlviv	-0.648***	-0.604***	-0.617***
	[0.013]	[0.013]	[0.013]
dumodesa	-0.585***	2.154***	-0.538***
	[0.014]	[0.329]	[0.013]
dumrivne	-1.071***	-1.024***	-0.928***
	[0.018]	[0.015]	[0.013]
dumzhytomyr	-1.005***	-1.001***	-0.802***
	[0.018]	[0.020]	[0.017]
Constant	6.337***	8.038***	6.912***
	[0.448]	[0.308]	[0.531]
Observations	3408	3408	3408
R-squared	0.8	0.8	0.79

Standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Special point of interest in SAR model is the coefficient near spatial lag. If it is insignificant, we get OLS. As was shown by LM test for spatial lag, spatial lag has to be included into the model and it turned out to be significant.

According to the results if housing price increases by 1% housing price in administrative center of the region (oblast) that share common border with region (oblast), where increase in housing price was observed, increases by 0.254%.

If prices of apartments in neighbor cities increase (neighbor cities are those for which positive weight is given), if to be more precise, their weighted average increases by 1%, housing price in administrative center of the region (oblast) increases by 0.368%. Since cities with similar number of inhabitants are given higher weights by construction of weighting matrix based on the demographic characteristics, rise of prices in similar cities (criterion of similarity – population) has higher influence on prices of apartments in the city of interest.

If the weighted average of housing prices in administrative centers of the region (oblast) goes up by 1%, prices in the city of interest go up by 0.43%. Similarly to the previous example, rise of prices in neighbor cities in terms of geography has higher effect (by the construction of weighting matrix).

On the other hand, standard errors are quite large if to compare with previously estimated model (spatial error model) because spatial error correlation was not taken into account. Thus, all three models are quite similar according to their explanatory power and have the same explanatory power as OLS.

The next model should take into account all spatial aspects and perform better than previous models.

- Spatially autoregressive model with spatially correlated error terms:

Table 5.8: Estimation results from spatially autoregressive model with spatially correlated error terms

lnprice	Weighting matrix based on the inverse of distance	Weighting matrix based oncontiguity	Weighting matrix based on the demographic characteristics
wlnprice	**0.412*** [0.045]**	**0.235*** [0.030]**	**0.360*** [0.052]**
floor	-0.027*** [0.006]	-0.030*** [0.006]	-0.023*** [0.006]
area	0.019*** [0.002]	0.025*** [0.001]	0.021*** [0.002]
repair	0.069*** [0.006]	0.071*** [0.007]	0.070*** [0.006]
material	0.050*** [0.006]	0.054*** [0.006]	0.048*** [0.006]
areasq	-0.000*** [0.000]	-0.000*** [0.000]	-0.000*** [0.000]
dumcherkasy	-1.127*** [0.020]	-1.230*** [0.034]	-0.969*** [0.014]
dumdnipropetrovsk	-0.796*** [0.014]	-0.763*** [0.013]	-0.811*** [0.017]
dumkharkiv	-0.937*** [0.015]	-0.917*** [0.014]	-0.914*** [0.014]
dumlviv	-0.646*** [0.013]	-0.604*** [0.013]	-0.618*** [0.012]
dumodesa	-0.586*** [0.014]	-0.06 [0.389]	-0.542*** [0.013]
dumrivne	-1.066*** [0.018]	-1.019*** [0.015]	-0.928*** [0.013]
dumzhytomyr	-0.998*** [0.020]	-0.991*** [0.020]	-0.803*** [0.017]
Constant	10.583*** [0.711]	10.243*** [0.365]	10.388*** [0.755]
Observations	3408	3408	3408
R-squared	0.83	0.96	0.86

Standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

As it can be inferred from the table above the coefficient near spatial lag is significant in all specifications. The coefficients near spatial lag decreased in magnitude comparatively with spatially autoregressive model (all other coefficients also decreased), but their standard error decreased as well. Moreover, according to our results model with weighting matrix based on contiguity performed better than other two models in terms of explanatory power.

In this model we assume that changes in housing prices in one city directly influences changes in housing prices in the other city, for instance, because if individuals observe increase in housing prices in some city, they anticipate raise of housing prices in their native city, especially when two cities are neighbors (either geographically – if we use contiguity or inverse distance matrices, or demographically – if we use weighting matrix based on demographic characteristics). In addition, neighbor cities may have similar economic conditions, educational opportunities or other characteristics, which are not always possible to measure and to include into the regression directly, therefore error terms exhibit spatial correlation.

Finally, spatial dependency of housing prices on the level of one city – Kyiv – is considered. Kyiv consists of 10 districts, which are assumed to be relatively homogeneous. Only weighting matrix based on contiguity is applied to explore spatial correlation between housing prices, since data about population characteristics is included directly into the model.

A Lagrange Multiplier test for spatial error and spatial correlation (presented in Appendix C) indicates the need of using spatially autoregressive model with spatially correlated error terms; however, again for the sake of comparison we present the results from running OLS, spatial error model and spatial autoregressive model. Though dummies for districts were found to be significant, they were not included into the model due to the high degree of collinearity with neighbor variables represented by the level of migration per 1000 inhabitants, ratio of crime rate in the district to the population in the district, ratio of workplaces to the total number of the population and ratio of number of students at schools to the total number of children in the age 6-18 in the district.

Table 5.9: Regression results for Kyiv housing market

lnprice	OLS	SAR	SEM	SARMA
area	0.037***	0.032***	0.037***	0.026***
	[0.003]	[0.003]	[0.003]	[0.001]
areasq	-0.000***	-0.000***	-0.000***	-0.001***
	[0.000]	[0.000]	[0.000]	[0.000]
floor	-0.028***	-0.030***	-0.028***	-0.031***
	[0.010]	[0.010]	[0.010]	[0.010]
mater	0.039**	0.040**	0.039**	0.039**
	[0.016]	[0.016]	[0.016]	[0.016]
repair	0.147***	0.151***	0.147***	0.147***
	[0.017]	[0.017]	[0.017]	[0.017]
workplaces	0.003***	0.003***	0.003***	0.003***
	[0.000]	[0.000]	[0.000]	[0.000]
crime	-7.297	-16.494**	-7.338	-28.830***
	[7.484]	[8.068]	[7.487]	[8.166]
migr	-0.007***	-0.010***	-0.007**	-0.013***
	[0.002]	[0.002]	[0.002]	[0.002]
pupiltototal	0.104**	0.103**	0.104**	0.143***
	[0.041]	[0.041]	[0.041]	[0.042]
wlnprice		**0.109*** **		**0.284*** **
		[0.036]		[0.048]
Constant	10.161***	9.081***	10.202***	8.109***
	[0.071]	[0.362]	[0.071]	[0.584]
Observations	1690	1690	1690	1690
R-squared	0.73	0.73	0.73	0.76

Standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

According to the regression results among the structural characteristics repair is valued the most – an apartment with repair will be sold 14.5 % more expensive than the one without it. Such characteristics as material, floor, where an apartment is located, area and area squared are significant and have the signs, which we expected, however, area squared does not seem to be significant economically (it is significant statistically, but very small in magnitude).

Neighbor variables also affect prices of apartments. Since the ratio of the number of workplaces in the district to the total population in the district is significant and has positive sign, we can state that mainly inhabitants of Kyiv do not like to live in bedroom communities and value more districts with larger number of working places.

Change in the population due to migration (measured per 1000 inhabitants) has negative sign, which indicates that if apartments in some district have lower prices than everywhere around, more newcomers are arriving to that district.

The effect of the number of crimes on apartment prices is negative; and it is quite large. It seems that citizens of Kyiv value safety a lot.

Ratio of total number of children at the age 6-18 to the number of students at schools shows that some districts have much more students at school than total number of children in that districts, while other districts have just opposite situation. This ratio measures quality of schools in the district and adequacy of the number of schools in the district given the number of children in the district. Good schools in the district add value to the apartments in the same district.

Estimation of spatial models indicates that prices of apartments in different districts are interconnected. According to the results of spatially autoregressive model with spatially correlated error terms (SARMA) – increase in weighted average of housing prices in the districts, which border the district of interest, by 1% leads to an increase of prices of apartments in the district of interest by

0.284% keeping all other things constant. This number is quite similar to the one obtained from the analysis of spatial correlation in housing prices between Ukrainian cities using contiguity matrix – we got 0.235.

The next section summarizes all findings for Ukrainian housing market as well as Kyiv city market.

Chapter 6

CONCLUSIONS

Apartments are heterogeneous goods, which are valued not only according their structural characteristics (number of rooms, total living area, repair), but also by location and neighbor characteristics (unemployment rate, ecological situation, quality of schools etc.). Similar apartments located in different places may be valued differently and spatial heterogeneity is observed. At the same time prices of apartments located in neighbor cities are correlated. Cities can be neighbors not only in terms of geography, but also on the basis of demographic and socio-economic characteristics. Price in neighbor cities can be directly affected by each other and in this case spatial lag of the dependent variable is included into the model. Spatial lag is nothing else but weighted average of the housing prices in the neighbor cities. Another case is correlation in error terms, which is actually statistical nuisance due to the omitted neighbor variable, which could influence housing prices in the neighbor cities. Finally, one model can contain both spatial lag and spatial error.

In the thesis I explored housing prices in eight Ukrainian cities (Kyiv, Kharkiv, Cherkasy, Odesa, Rivne, Lviv, Zhytomyr and Dnipropetrovsk), which are oblast centers. In addition the market of one city (Kyiv) was analyzed. According to the estimation results spatial heterogeneity is present in Ukrainian real estate market. What is more, not only the intercepts differ, but also slopes. Area and repair are valued higher in Kyiv than in any other city used in the regression analysis, while floor is insignificant for Kyiv and has negative values in Kharkiv and Dnipropetrovsk and inhabitants in the latter cities are willing to pay higher prices for apartments located in houses made of bricks than inhabitants of Kyiv. As it was expected a priori apartments located on the first/last floor are valued less due to the danger of burglary or problems with the roof. Larger area, recent repair and brick as a type of material for the house, where an apartment is located, increase the price of the apartment.

Spatial dependency is also a feature of housing market in Ukraine. We explored spatial dependency between prices of similar apartments located in different cities and also different districts of one city (Kyiv). The criterion for similarity was total living area of an apartment. In both cases robust LM tests showed the need for inclusion into the model spatial lag and spatially correlated error terms. Need for the spatial lag indicates that housing prices in housing prices in one city directly affect housing prices in the neighbor city, while spatially correlated error terms result from omitted neighbor characteristics.

In practice spatially autoregressive model with spatially correlated error terms outperformed other models in terms of explanatory power. If to compare spatial models with OLS, which was estimated mainly as a benchmark, in most of the cases the sign and the level of significance of the coefficients didn't change, but their magnitude did. On average the coefficients became smaller.

Three different weighting matrices were taken in order to capture spatial dependency between housing prices in different cities: weighting matrix based on contiguity, on the inverse of distance and on the inverse of difference in the number of inhabitants in the city. In other words, the first two matrices are based on geographical notion of neighbors, while the last – on the demographic characteristics. The results showed that prices are correlated in geographically neighbor cities and in cities with similar demographic characteristics.

The findings may be of interest for real estate experts and economists, because the model is designed to help predict housing prices in different regions if we know housing characteristics.

Due to the underdevelopment of equity market in Ukraine citizens often invest money into apartments, therefore, individuals who are going to buy or sell a house and investors may be also an interested party in this research because they can determine the place, where they want to buy an apartment.

Since implicit prices of apartment attributes were calculated, building companies may take it into consideration when planning the project of the house. Apart from higher prices that they can charge for an apartment built according to the preferences of inhabitants, consumers' satisfaction from quality of apartments will rise.

For the case of districts of Kyiv the implicit price of school quality, implicit cost of crime rate and preferences of inhabitances about the presence of working places in the district were evaluated. Policy makers can use these results for urban planning and development in order to allocate taxes paid by the inhabitants of districts in the most efficient and fair way (taking into account possible increase in social welfare).

BIBLIOGRAPHY

Anselin, Luc. 1999. *Spatial Econometrics.* University of Texas at Dallas Working Paper.

Baumont, Catherine. 1999. *Spatial Effects in Housing Price Models. Do Housing Prices Capitalize Urban Development Policies in the Agglomeration of Dijon (1999)?* CNRS UMR 5118

Besner, Claude. *Spatial Autoregressive Specification with a Comparable Sales Weighting Scheme.* JRER, Vol. 24, No. 2 (2002), pp. 193 – 210.

Bitter, Chris, Gordon Mulligen and Sandy Dall'erba. 2006. *Incorporating Spatial Variation in Housing Attribute Prices: A Comparison of Geographically Weighted Regression and the Spatial Expansion Method.* MPRA Paper # 1379.

Brady, Ryan. 2007. *Measuring the Persistence of Spatial Autocorrelation: How Long Does the Spatial Connection between Housing Markets Last?* Accessed on-line at: ideas.repec.org/p/usn/usnawp/19.html

Brunsdon, Chris, Steward Fortheringham and Martin Charlton. *Geographically Weighted Regression – Modelling Spatial Non-Stationarity.* The Statistician, Vol. 47, No.3 (1998), pp. 431- 443

Case, Bradford, John Clapp, Robin Dubin and Mauricio Rodriguez. 2003. *Modeling Spatial and Temporal House Price Patterns: A Comparison of Four Models.* Accessed on-line at: business.uconn.edu

Chekmezova, Olena. 2007. *Do We Pay for Clean Air: Hedonic Price Analysis of Housing Market in Kyiv.* 2007. EERC MA thesis.

Clapp, John. 2002. *A Semi Parametric Method for Estimating Local Price Indices.* Accessed on-line at: http://www.sba.uconn.edu/users/johnc/currentresearch.html

Darmofal, David. 2006. *Spatial Econometrics and Political Science.* Accessed on-line at: polmeth.wustl.edu/retrieve.php?id=575

Dubin Robin. *Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms.* The Review of Economics and Statistics, Vol. 1, No. 3. (Aug., 1988), pp. 466-474.

Fotheringham, Steward, Martin Charlton and Chris Brundson. 2000. *Geographically Weighted Regression.* Accessed on-line at: ncg.nuim.ie/GWR

Gao, Xiaolu, Yasushi Asami and Chang Yo Chung. 2002. *An*

Empirical Evaluation of Hedonic Regression Models. CSIS Discussion Paper No. 46.

Karlsson, Vifill. *Modern Industrial Structure and Development of House Price Spatial Disparity: A General Case for Iceland – a Large, but Thinly Populated European Country.* Accessed on-line at: www.bifrost.is/Files/Skra_0022952.pdf

Kelejian, H. and I. Prucha. 1997. *Estimation of Spatial Regression Model with Autocorrelated Disturbances.* Working paper.

Long F., A. Paez and S. Farber. 2007. *Spatial Effects in Hedonic Price Estimation: a Case Study in the City of Toronto.* Accessed on-line at: sciwebserver.science.mcmaster.ca/cspa/papers/index.html

Mavrodiy, Antonina. 2005. *Factor Analysis of Real Estate Prices.* EERC MA thesis

McMillen, Daniel. *The Return of Centralization to Chicago: Using Repeat Sales to Identify Changes in House Price Distance Gradients.* Regional Science and Urban Economics, Vol. 33, 2003, pp. 287 - 304

Ord, Keith. *Estimation Methods for Models of Spatial Interaction.* Journal of the American Statistical Association. Vol. 70, No. 349. (Mar, 1975), pp. 120 – 126.

Pace, R., Barry K. and C. Sirmans. *Spatial Statistics and Real Estate.* Journal of Real Estate Finance and Economics. Vol. 17, 1998, pp. 5-13.

Rosen, Sherwin. *Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition.* The Journal of Political Economy. Vol. 82, No. 1 (Jan. – Feb., 1974), pp. 34 – 55.

Sioma, Dmytro. 2006. *Rental Housing Price Differentials and the Cost of Commuting: Evidence from Kyiv.* EERC MA thesis.

Sommerville, Tsur and John Ries. 2004. *School Quality and Residential Property Values: Evidence from Vancouver Rezoning.* Center for Urban Economics and Real Estate Working Paper 2004-01.

Tsutsumi, Morito, Yasushi Yoshida, Hajime Seya and Yuichiro Kawaguchi. 2005. *Spatial Analysis of Tokyo Apartment Market.* Accessed on-line at: fp.paceprojects.f9.co.uk/**Tsutsumi**.pdf

Wallace, Nancy. *Hedonic-Based Price Indexes for Housing: Theory, Estimation, and Index Construction.* FRBSF Economic Review, No. 3, 1996

APPENDIX A

THREE STEP PROCEDURE ESTIMATION

Eight assumptions are needed to estimate the model:

(1) Elements of weighting matrices, which are located on the main diagonal $\{w_{ii}\}$ and $\{m_{ii}\}$ are equal to zero.
(2) $(I - \lambda W_n)$ and $(I - \varrho M_n)$ are nonsingular
(3) Matrix of independent variables X has a full column rank and its elements are bounded uniformly in absolute value.
(4) Sums of elements in rows and in columns in matrices W_n, M_n, $(I - \lambda W_n)^{-1}$ and $(I - \varrho M_n)^{-1}$ are bounded uniformly in absolute value.
(5) $\varepsilon \sim iid(0, \sigma^2_\varepsilon I)$ and $0<\sigma^2_\varepsilon < c$ and $c < \infty$. Error terms ε have finite fourth moments.
(6) If we denote a matrix of instruments used for estimation of the model by H_n, then it has full column rank and consists of the linearly independent columns of $(X_n, W_n X_n, W_n^2 X_n, M_n X_n, M_n W_n X_n \ldots)$
(7) The instruments have the following properties

$$Q_{HH} = \lim_{n\to\infty} n^{-1} H_n' H_n$$

Q_{HH} – finite and nonsingular

$$Q_{HZ} = p\lim_{n\to\infty} n^{-1} H_n' Z_n$$

$$Q_{MHZ} = p\lim_{n\to\infty} H_n' M_n Z_n$$

Q_{HZ} and Q_{HMZ} – finite with full column rank

$Q_{HZ} - \rho Q_{HMZ} = p\lim_{n\to\infty} n^{-1} H_n{}'(I-\rho M_n)Z_n$ has full column rank as long as $|\rho|<1$

$\Phi = \lim_{n\to\infty} H_n{}'(I-\rho M_n)^{-1}(I-\rho M_n{}')^{-1} H_n$ is finite and nonsingular as long as $|\rho|<1$

(8) The smallest eigenvalue of $\Gamma_n{}'\Gamma_n > 0$

$$\Gamma_n = \frac{1}{n}\begin{pmatrix} 2E(u_n{}'u_n) & -E(\bar{u}_n{}'\bar{u}_n) & 1 \\ 2E(\bar{\bar{u}}_n{}'\bar{u}) & -E(\bar{\bar{u}}_n{}'\bar{\bar{u}}_n) & tr(M_n{}'M_n) \\ E(u_n{}'\bar{\bar{u}}_n + \bar{u}_n{}'\bar{u}_n) & -E(\bar{u}_n{}'\bar{\bar{u}}_n) & 0 \end{pmatrix}$$

where $\bar{u}_n = M_n u_n$ *and* $\bar{\bar{u}}_n = M_n \bar{u}_n = M_n^2 u_n$

First, the model will be rewritten as:

$Y_n = Z_n\delta + u_n$

$u_n = \lambda M_n u_n + \varepsilon_n$

$Z_n = (X_n, W_n Y_n)$ and $\delta = (\beta', \rho')'$

After Cochorane-Orcutt transformation:

$Y_{n*} = Z_{n*}\delta + \varepsilon_n$

and $Y_{n*} = Y_n - \lambda M_n Y_n, \quad Z_{n*} = Z_n - \lambda M_n Z_n$

Step 1.

Estimate $\delta_n = (\hat{Z}_n{}'\hat{Z}_n)^{-1}\hat{Z}_n{}'Y_n$

$$\hat{Z}_n = P_{Hn}Z_n = (X_n, \overline{W_nY_n}), \quad \overline{W_nY_n} = P_{Hn}W_nY_n \quad and \quad P_{Hn} = H_n(H_n{}'H_n)^{-1}H_n{}'$$

<u>**Step 2.**</u>

Let's denote the i-th element of u_n by $u_{i,n}$, $\bar{u}_{i,n}$ *as an element of* $\bar{u}_n = M_n u_n$ *and* $\bar{\bar{u}}_{i,n}$ *of* $\bar{\bar{u}}_n = M_n^2 u_n$.

Then moments will be estimated:

$$E(\frac{1}{n}\varepsilon'\varepsilon) = \sigma^2 \qquad E(\frac{1}{n}\varepsilon'M'M\varepsilon) = \sigma^2\frac{tr(M'M)}{n} \qquad E(\frac{1}{n}\varepsilon'M'\varepsilon) = 0$$

The moment conditions can be rewritten in terms of u:

$$E(\frac{1}{n}u'(I-\lambda M)'(I-\lambda M)u) = \sigma^2$$

$$E(\frac{1}{n}u'(I-\lambda M)'M'M(I-\lambda M)u) = \frac{\sigma^2}{n}tr(M'M)$$

$$E(\frac{1}{n}u'(I-\lambda M)'M'(I-\lambda M)u) = 0$$

Next the above three terms will be rearranged in the following way:

- the first term

$$\frac{1}{n}E(u'u) - \lambda(\frac{2}{n}E(u'Mu)) + \lambda^2(\frac{1}{n}E(u'M'Mu)) = \sigma^2$$

$$=> [\frac{2}{n}E(u'\bar{\bar{u}}) \quad \frac{-1}{n}E(\bar{u}'\bar{u}) \quad 1]\times[\lambda \quad \lambda^2 \quad \sigma^2]-\frac{1}{n}E(u'u)=0$$

- the second term

$$[\frac{2}{n}E(\bar{\bar{u}}'\bar{u})] \quad \frac{1}{n}E(\bar{\bar{u}}'\bar{\bar{u}}) \quad \frac{1}{n}tr(M'M)]\times[\lambda \quad \lambda^2 \quad \sigma^2]-\frac{1}{n}E(\bar{u}'\bar{u})=0$$

- the third term

$$[\frac{1}{n}E(u'\bar{\bar{u}}+\bar{u}'\bar{u}) \quad \frac{-1}{n}E(\bar{u}'\bar{\bar{u}}) \quad 0]\times[\lambda \quad \lambda^2 \quad \sigma^2]-\frac{1}{n}E(u'\bar{u})=0$$

$$\alpha=(\lambda,\lambda^2,\sigma_\varepsilon^2)' \quad and \quad \gamma=n^{-1}(E(u_n{}'u_n),E(\bar{u}_n{}'\bar{u}_n),E(u_n{}'\bar{u}_n))'$$

$\Gamma_n\alpha=\gamma_n$ or in estimated form $g_n=G_n\alpha+v_n$

Γ is specified in Assumption 8

$$\alpha_n=G_n^{-1}g_n$$

Having replaced population moments by sample moments, estimators of λ and σ_ε^2 can be obtained from the non-linear least square estimator of α_n

Step 3.

When we know estimator of λ we are able to calculate feasible GS2SLS estimator of σ:

$$\sigma_{F,n}=[Z_{n*}(\lambda_n)'Z_{n*}(\lambda_n)]^{-1}Z_{n*}(\lambda_n)'Y_{n*}(\lambda_n)$$

Where

$$Z_{n*}(\lambda_n) = P_{Hn} Z_{n*}(\lambda_n), \quad Z_{n*}(\lambda_n) = Z_n - \lambda_n M_n Z_n, \quad Y_{n*}(\lambda_n) = Y_n - \lambda_n M_n Y_n$$

$$Z_{n*}(\lambda_n) = (X_n - \lambda_n M_n X_n, \quad W_n Y_n - \lambda_n M_n W_n Y_n) \quad and \quad W_n Y_n - \lambda_n M_n W_n Y_n = P_{Hn}(W_n Y - \lambda_n M_n W_n Y_n)$$

APPENDIX B

Table B.1: Results of tests for spatial dependence in OLS specification for Ukrainian housing market

Type of test	Value	Critical values	Decision rule
	Weighting matrix based on contiguity		
LM_{error}	0.4462		Do not reject null
LM_{lag}	15.3678		Reject null
LM_{error}^{robust}	56.7770	P=0.05: 6.635 P=0.01: 3.841	Reject null
LM_{lag}^{robust}	33.9368		Reject null
	Weighting matrix based on the inverse of distance		
LM_{error}	0.1770		Do not reject null
LM_{lag}	16.0538		Reject null
LM_{error}^{robust}	69.8800	P=0.05: 6.635 P=0.01: 3.841	Reject null
LM_{lag}^{robust}	39.8638		Reject null
	Weighting matrix based on demographic characteristics		
LM_{error}	5.7952		Reject null
LM_{lag}	0.1400		Do not reject null
LM_{error}^{robust}	17.0444	P=0.05: 6.635 P=0.01: 3.841	Reject null
LM_{lag}^{robust}	21.0329		Reject null

APPENDIX C

Table C.1: Results of tests for spatial dependence in OLS specification (model with interactions) for Kyiv housing market

Type of test	Value	Critical values	Decision rule
LM_{error}	176.03		Reject null
LM_{lag}	94.3867		Reject null
LM_{error}^{robust}	147.88	P=0.05: 6.635 P=0.01: 3.841	Reject null
LM_{lag}^{robust}	1.9120		Do not reject null