DO WE PAY FOR CLEAN AIR:
HEDONIC PRICE ANALYSIS OF
HOUSING MARKET IN KYIV

by

Olena Chekmezova

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

Master of Arts in Economics

National University "Kyiv-Mohyla Academy"
Economics Education and Research Consortium
Master's Program in Economics

2007

Approved by ___________________________________
Mr. Serhiy Korablin (Head of the State Examination Committee)

____

____

____

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

Date ___________________________________

National University "Kyiv-Mohyla Academy"

Abstract

DO WE PAY FOR CLEAN AIR:
HEDONIC PRICE ANALYSIS OF
HOUSING MARKET IN KYIV

by Olena Chekmezova

Head of the State Examination Committee: Mr. Serhiy Korablin,
Economist, National Bank of Ukraine

This paper provides estimation of marginal willingness to pay for the air quality improvement using hedonic price analysis. It is conducted for housing market in Kyiv and is motivated by the high air pollution in Kyiv on the one hand, and lack of applications of hedonic price techniques for valuation environmental amenities in transition countries on the other hand. This study utilized data on rental and sales housing markets with combination of the data on four air pollutants: particulate matter (PM), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), and carbon monoxide (CO). It appears that residents of Kyiv are concerned more about pollution that comes from stationary sources comparing to pollution that comes from mobile emission sources. It is indicated by their positive willingness to pay for improvement in SO_2 concentrations. This relationship is supported in the analysis of both rental and sales housing market data, which confirms the idea that residents of Kyiv do value clean air.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express the deepest appreciation to my thesis adviser, Professor Denys Nizalov, for his supervision of this research and giving me invaluable comments and suggestions. He was always there to meet and talk about my ideas, read my numerous drafts, and provide encouragement.

My special thanks goes to Mykhaylo Salnykov for giving me inspiration for the topic of my thesis and giving me many practical advices at the initial stages of this paper.

I am grateful to Tom Coupé, Olena Nizalova, Olesia Verchenko, Oleksandr Shepotylo, Pavlo Prokopovych, Hanna Vakhitova and all other professors of EERC MA Program in Economics who reviewed my paper and gave me thoughtful comments and recommendations.

I also thank my EERC friends, Dzmitry Sidarau for helping me to process and analyse the data and for having confidence in me, and Olena Senyuta for patient listening to my complaints and frustrations.

Chapter 1

INTRODUCTION

World Development Indicator 2005 continues to show that middle-income countries have high level of air pollution comparing with high-income or very poor countries. In early 1990s the relationship between country's wellbeing and environmental deterioration resulted in a hypothesis that middle-income countries place less value on the quality of environmental goods than rich countries. This relationship is known as environmental Kuznets curve and was popularized by the *World Development Report* (1992). Since Ukraine, according to the *World Bank* classification, belongs to the lower-middle-income country, it can be hypothesized (on individual level) that people in Ukraine do not pay significant attention to air quality comparing to people in high-income countries.

Therefore, the main purpose of this research is (i) to assess the economic value that people in Kyiv place on clean air using hedonic price techniques, specifically to derive marginal willingness to pay for the air quality improvement, and (ii) to compare the obtained results with the respective estimates from other countries.

The non-market nature of environmental benefits comprises the main difficulty in deriving and eliciting their economic value. One method to assess the value placed on environmental goods, including clean air, is the hedonic price analysis. The underlying assumption of this method is that the price of a quality-differentiated marketed good, such as housing, can be represented as a function of different characteristics, including the environmental characteristics (the idea that people usually derive utility not from the good itself but rather from its characteristics and properties was first expressed by Lancaster, 1966). Thus, the houses with the same non-environmental characteristics will cost more if they are located in the area with less air pollution (assuming that people are willing to pay a positive amount to avoid pollution). This price difference provides a base for estimation of the value of clean air.

There are many studies that extensively use hedonic price analysis to address the relationship between housing value and air pollution. One of the most oft-cited papers is by *Smith* and *Huang* (1995), who provides a thorough review and analysis of over 50 studies in the US cites conducted between 1967 and 1988. Most of these studies reinforce the idea that air pollution has an impact on housing value. More recent studies (Chattopadhyay, 1999, Ekeland et al., 2002, Kim et al., 2003) estimate willingness to pay for non-marginal changes in air quality and address different econometric problems. Although hedonic price analysis serves as a well-developed method of applied econometrics and has been used in many developed countries to estimate the welfare effects from reducing air pollution, research on developing countries and countries in transition is very scarce. Lack of studies in transition countries and non-existence of them on Ukraine provoke conducting this research.

This study utilized data on rental and sales housing markets with combination of the data on four air pollutants: particulate matter (PM), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), and carbon monoxide (CO). It implements principal component analysis to analyze data on air pollution and uses hedonic price techniques for analysis of housing market in Kyiv.

This research has important policy implications. Knowing how much value people place on air quality, policy makers could compare costs associated with the reduction in air pollution and benefits from improving the air quality while implementing environmental policies. Moreover, real estate specialists will benefit from knowing the price of environmental component in the value of houses.

The remainder of the thesis is organized as follows. Chapter 2 includes the discussion of the theoretical underpinnings of the hedonic price analysis and reviews the current literature that demonstrates the air pollution - housing relationship. Next two chapters provide description of the data and estimation framework . Chapter 5 presents the estimation results. In conclusion, we provide discussion of the results and policy implications.

Chapter 2

LITERATURE REVIEW

Current study conducts an assessment of value of air pollution in the city of Kyiv using the hedonic price technique. Therefore, we first present the theoretical framework including the discussion of econometric issues that arise while implementing the method; and then turn to the overview of the literature that demonstrates the air pollution - housing market relationship.

Theoretical framework for the hedonic price analysis was described by *Rosen* (1974). He considers the competitive housing market with home owners maximizing their profits and buyers maximizing their utility. We can think about houses as differentiated commodities that differ among each other in structural characteristics (total floor area of the apartment, number of rooms, housing age etc.), characteristics of the neighborhood (proximity to the green spaces, proximity to the metro station etc), and characteristics that describe environmental conditions (concentrations of air pollutants). Rosen's model assumes that consumers derive utility from all characteristics of the housing, while home owners (producers in Rosen's model) face costs to produce the commodity with different characteristics. The result of interactions of these two groups can be described as hedonic price schedule.

Assuming that market is in equilibrium and denoting by $\vec{x}_i = (x_{i1},...,x_{in})$ the vector of *n* characteristics of differentiated commodity *i*, i.e. housing, the price of this housing can be written as $p_i = p(\vec{x}_i)$. The last expression is called the hedonic price function. Rosen argues that hedonic price function is nonlinear since a linear function can only arise if separate characteristics of the commodity can vary independently of the good itself. The hedonic price function can be derived formally by analyzing behavior of buyers and home owners. The consumer's problem can be stated as:

$$\max U(y, \vec{x}_i, \vec{\alpha})$$

$$\text{s.t. } y + p(\vec{x}_i) = C,$$

where y is the value of all other goods except housing, $\vec{\alpha}$ – vector of parameters that describe consumer's preferences, and C is the amount of income available to the consumer. While writing this problem, we assume that consumers buy only one housing in the competitive market. Solving this maximization problem, we can find the optimal condition for the characteristic of our interest x_{ij} (environmental one, that is ambient level of pollutant): $\frac{\partial U(\vec{y}, \vec{x}_i, \vec{\alpha}) / \partial x_{ij}}{\partial U(\vec{y}, \vec{x}_i, \vec{\alpha}) / \partial \vec{y}} = \frac{\partial p(\vec{x}_i)}{\partial x_{ij}}$. We can interpret this first order condition as marginal willingness to pay for one additional unit of the characteristic x_{ij} must equal to marginal cost of this characteristic while holding all other things constant (Freeman, 2003). $\frac{\partial p(\vec{x}_i)}{\partial x_{ij}}$ is the implicit marginal price of the characteristic x_{ij}. Therefore, taking the partial derivative of the hedonic price function with respect to a characteristic x_{ij}, we can calculate the implicit marginal price of this characteristic. Based on hedonic price function, the welfare effects can be estimated. Considering the improvement in air quality, the gain to the affected house can be described as $WTP = p(\vec{x}_i^{\,*}) - p(\vec{x}_i)$ with $\vec{x}_i^{\,*}$ being the new vector of housing characteristics and *WTP* being the willingness to pay (Haab, 2002).

Estimating hedonic price function, significant attention should be paid to the model specification and possible econometric and practical problems. Thus, we review three econometric issues that may arise while implementing this method:

(i) specification form (choice of the dependent and independent variables); (ii) functional form; and (iii) market segmentation.

Talking about the specification issues, we should be concerned about the source of data on housing prices and set of included variables.

Selecting appropriate independent variables is not an easy task since the theory does not give any guidelines concerning this issue. Many studies try to control for as much variables as possible in order to estimate the effect of environmental characteristics on housing prices more precisely. But including a lot of housing characteristics into the hedonic price function results in multicollinearity problem among housing characteristics and possible endogeneity problems. These problems are even more amplified when we deal with the data on developed countries: housing markets in most developed countries are organized in such a way that higher-quality houses are often allocated in areas with better environmental conditions, such as in the proximity to the green spaces (Harrison and Rubinfeld, 1978). Since housing market structure of Kyiv and also other large cities of Ukraine was inherited from the Soviet Union times, such sorting might not be so severe in Kyiv at this moment. Thus, conducting research on transitional country may help to partially overcome the above mentioned problems.

There is a study based on Kyiv's housing market by Mavrodiy (2005), which analyzes macro and micro determinants of real estate prices. Based on advertisement data from "Aviso" newspaper from 2004 and 2005 years, the author estimates two different regressions from each data set including 13 variables as explanatory variables in regression analysis. She concludes that there is a significant impact on housing prices of the following housing characteristics: number of rooms, total floor area of the apartment (significant only for 2004 data set), wall material of the building (significant only for 2005 data set), the height of ceiling, availability of furniture in the apartment (significant only for 2004 data set) location (district), and location relative to the nearest metro station. These results can be used as a first insight on what variables should be included in the hedonic price equation as explanatory variables.

Now we focus on the functional form. The theory does not place any restriction on the functional form of the hedonic price equation because the last one is derived based on the interactions between utility maximizing buyers and profit maximizing home owners (Freeman, 2003). This explains the variability of functional forms in the empirical literature. Early studies used primary simple forms: linear, semi-log, or log-log. Later the most general and flexible functional form was proposed by *Halvorsen* and *Pollakowski* (1981):

$$p_i(\lambda) = \alpha_0 + \sum_{j=1}^{n} \alpha_j x_{ij}(\mu) + \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} \beta_{jk} x_{ij}(\theta) x_{ik}(\theta) + \varepsilon_i \ ,$$

where ε_i is independently and identically distributed error term; α's and β's are associated parameters; and $p_i(\lambda) = \left(p_i^{\lambda} - 1\right)/\lambda$ for $\lambda \neq 0$ and $\ln p_i$ for $\lambda = 0$. Although flexible form seems to be very attractive, it amplifies the already significant multicollinearity problem of characteristics because of the presence of interaction terms. So, there is always a tradeoff between increasing multicollinearity among housing characteristics and flexibility in the functional form.

Cropper et al. (1988) provides some guidelines in making decision about functional form. Using Multiple Listing data on housing characteristics in Baltimore City, the authors simulate housing market equilibrium and based on specified utility function for the households compute true marginal implicit prices of housing characteristics. Based on equilibrium prices they are able to estimate six different functional forms of hedonic price function and compare the calculated marginal implicit prices with the true ones. They conclude that if all housing characteristics were included in the hedonic price function, the unrestricted Box-Cox model and linear Box-Cox model produce the most accurate results. If there is a problem of omitted variables or some variables were replaced by proxies, simpler models (linear, semi-log, linear Box-Cox) provide more accurate estimates of marginal implicit prices.

Based on the results found by *Cropper* et al. (1988), we can conclude that simple linear form outperforms the others since it is very likely to have a problem of omitted variables. But as stated above, linear function can only arise if separate characteristics of commodity can vary independently of the good itself (Rosen, 1974), which is impossible. Therefore, in this research we follow the procedure used by *Chattopadhyay* (1999) and estimate six different regressions by restricting parameters λ, μ, θ : (i) linear Box-Cox model with transformed dependent and independent variables using different parameters of transformation ($\beta_j = 0$); (ii) linear Box-Cox model with transformed dependent and independent variables using the same parameter ($\beta_j = 0$); (iii) linear Box-Cox model with transformed independent variables ($\lambda = 1$, $\beta_j = 0$); (iv) linear Box-Cox model with transformed dependent variable($\mu = 1$, $\beta_j = 0$); (v) semi-log model ($\lambda = 0$, $\mu = 1$, $\beta_j = 0$); and (vi) linear model ($\lambda = 1$, $\mu = 1$, $\beta_j = 0$).

The last estimation issue that we discuss concerns the extension of the housing market. All existing studies can be divided into three categories: (i) papers that treat national housing market as a single one, (ii) studies that define metropolitan area or labor market as a single housing market, and (iii) studies that adopt the notion of local market. While extending the definition of the housing market beyond the local one, the issue of the market segmentation should be carefully contemplated. The literature defines the market as segmented based on two criteria (Freeman, 2003). The first one is that buyers have different preferences or that the housing characteristics are different across submarkets. The second one is related to information or geographical barriers, which limit the mobility of buyers preventing them from purchases of houses in the whole area. If the market under analysis is actually segmented, separate hedonic price functions should be estimated for each submarket (Haab, 2002). Although we are not dealing with national market or labour market in the current research, we also test for the existence of housing market segmentation in Kyiv in order to ensure that we obtain valid marginal implicit prices while assuming that housing market in Kyiv is not segmented.

Now we briefly discuss the literature that demonstrates the air pollution-housing prices relationship. Good overviews of such studies can be found in *Smith* and *Huang* (1995), who present a meta-analysis of more than 50 studies in the US cites. More recent review is available in *Boyle* and *Kiel* (2001). But most of these studies are conducted based on developed housing markets while the problem looks more severe for developing countries. *Strukova* et al. (2006) in their recent paper stated that the annual average concentrations of PM2.5 (particulate matter less than 2.5 microns in diameter) in all 29 Ukrainian cities under investigation exceeded the maximum allowable concentrations. They found that health and mortality costs associated with air pollution in Ukraine are around 4% of GDP, which are very significant costs comparing to the EU countries, where respective costs are around 2% of GDP. We will not concentrate deeply on early studies, but rather discuss the recent research on developing and some developed countries.

Murty et al. (2003) provide estimation of hedonic price function for the rental housing market in Delhi and Kolkata, India. Based on data from the household survey in 2002 and monthly average concentrations of three pollutants (SPM, SO_2, and NO_x), they calculate marginal willingness to pay for the environmental characteristics. Estimating hedonic price functions, the authors use two approached: first they calculated hedonic prices for each city separately treating each of them as a separate market, and then they calculate marginal implicit prices using pooled data for both the cities. The conclusion is that two out of three pollutants had negative and significant impact on monthly rent independently of the approach used, while the coefficient on the third environmental variable have the wrong sign. The authors also calculate welfare gains from reducing air pollution in both cities arguing that such changes could be welfare enhancing.

Kim et al. (2003) significantly improve the methodology of estimating hedonic price function by explicitly incorporating the spatial effects into the model. They argue that if data are inherently spatial simple application of traditional hedonic approach may lead to the biased and inconsistent estimates of marginal implicit prices. The authors use spatial-lag and spatial-error models based on the linear hedonic price function in order to estimate the marginal improvements in air quality for Seoul housing market. They conclude that marginal willingness to pay for the decrease in mean SO_2 concentrations constitutes around 1.4% of mean housing price. Such negative association between pollution and housing prices in not, however, observed for nitrogen dioxide (NO_2).

Yusuf and *Resosudarmo* (2006) conduct hedonic price analysis to assess the value of clean air in Jakarta, Indonesia. They use Indonesian Family Life Survey for the period 1997-1998 as a source of housing data and data on the annual average concentrations of six different pollutants. As a dependent variable in the hedonic equation they use monthly house rental arguing that there is no significant difference between price or the rent of the house as a dependent variable. The explanatory variables are structural characteristics of the house, variables that reflected quality of the neighborhood, and ambient levels of pollutants. For each pollutant a separate hedonic price equation is estimated applying flexible functional form. The results of this study are mixed: the authors find a significant negative relationship between four out of six pollutants (SO_2, CO, THC, Pb) and housing rent, but no influence of other pollutants on the housing rental price. Therefore, most of the results do not support the idea that people in developing countries do not value clean area.

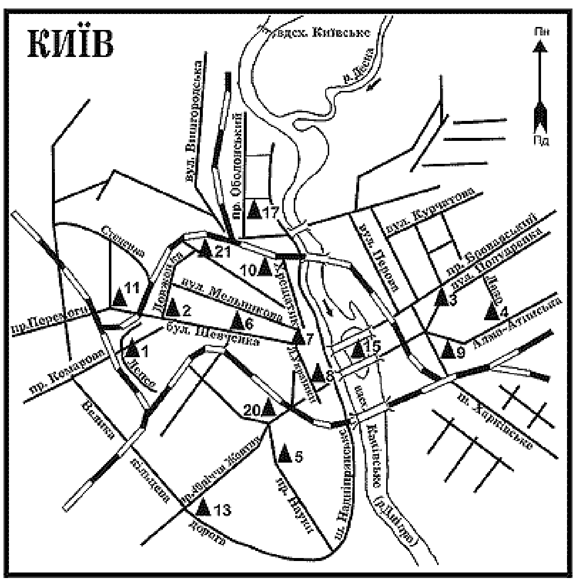
In conclusion, it should become clear that hedonic price analysis is well-developed technique that can be used to produce valuable policy recommendations. But still there are few studies that implement this method on the housing markets in the transition countries. So, our research will fill the existing gap in the literature and provide Ukrainian policymakers with useful recommendations.

DATA DESCRIPTION

The study utilizes the data on air pollution, which was kindly provided by the Central Geophysical Observatory (Ministry of Ukraine of Emergencies and Affairs of Population Protection from the Consequences of Chernobyl Catastrophe), and data on rental and sales hosing markets for 2005.

Air pollution is monitored four times per day at 16 monitoring stations (Figure 3.1), which are located in different parts of Kyiv.

Figure 3.1: Location of the monitoring stations



1. 6a Strazhesko st.
2. 8 Dovzhenko st.
3. 50 Popudrenko st.
4. 2 Lazo st.
5. 37 Nauky prosp.
6. Peremogy sq.
7. Besarabs'ka sq.
8. 29 Lesi Ukrayinky blvd.
9. 10a Kaunas'ka st.
10. 56/60 Mezhygirs'ka st.
11. 98/2 Peremogy prosp.
13. Expotsentr Ukrayiny
15. Hidropark
17. 14 Obolons'kyi prosp.
20. Moskovs'ka sq.
21. 5 Sklyarenko st.

Source: Central Geophysical Observatory www.cgo.org.ua

The data set contains the ambient levels of 20 contaminants for the year 2005: particulate matter (PM), sulfur dioxide (SO_2), soluble sulphates, carbon monoxide (CO), nitrogen dioxide (NO_2), nitrogen oxide (NO), hydrogen sulphide, phenol, hydrogen chloride, hydrogen chloride, ammonia, formaldehyde, and heavy metals. The primary air pollutants in Kyiv are particulate matter (PM), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), benzapilene, carbon monoxide (CO), phenol, and formaldehyde (Central Geophysical Observatory). Some of these pollutants come mainly from stationary sources (e.g. SO_2), while others – from mobile emission sources. Only 4 out of 20 pollutants are considerably reported by each monitoring station in Kyiv: PM, SO_2, NO_2, and CO. Thus, only these 4 pollutants are considered in the following analysis. Table 3.1 presents the year average levels of pollutants recorded by each monitoring stations.

Table 3.1: Statistics of Pollution in Kyiv, 2005

Station № and address	**PM annual concentr., mg/m^3**	**SO_2 annual concentr., mg/m^3**	**NO_2 annual concentr., mg/m^3**	**CO annual concentr., mg/m^3**
1. 6a Strazhesko st.	0.0856	0.0126	0.1177	1.6515
2. 8 Dovzhenko st.	0.0935	0.0141	0.1194	1.7350
3. 50 Popudrenko st.	0.0970	0.0106	0.1076	1.2673
4. 2 Lazo st.	0.1406	0.0118	0.1042	1.4405
5. 37 Nauky prosp.	0.0806	0.0103	0.0519	0.7311
6. Peremogy sq.	0.1026	0.0125	0.1023	2.4402
7. Besarabs'ka sq.	0.1514	0.0123	0.1326	2.8168
8. 29 Lesi Ukrayinky blvd.	0.1391	0.0132	0.1219	1.8399
9. 10a Kaunas'ka st.	0.1061	0.0132	0.1145	2.5852
10. 56/60 Mezhygirs'ka st.	0.1325	0.0130	0.1109	2.0148
11. 98/2 Peremogy prosp.	0.1672	0.0163	0.1202	2.1249
13. Expotsentr Ukrayiny	0.0730	0.0114	0.0501	0.8300
15. Hidropark	0.0881	0.0127	0.0703	1.0580
17. 14 Obolons'kyi prosp.	0.0991	0.0165	0.1244	1.3283
20. Moskovs'ka sq.	0.1194	0.0165	0.1093	2.6419
21. 5 Sklyarenko st.	0.1206	0.0154	0.0850	0.8749
Descriptive statistics of variables that describe environmental condition				
	PM	**SO_2**	**NO_2**	**CO**
Mean	0.1123	0.0133	0.1026	1.7113
Std. Dev.	0.0275	0.0020	0.0252	0.6806
Min	0.0730	0.0103	0.0501	0.7311
Max	0.1672	0.0165	0.1326	2.8168

During 2005 year annual average concentration of PM ranged from 0.0730 (Expotsentr Ukrayiny, station 13) to 0.1672 mg/m^3 (Peremogy prosp., station 11). Maximum observable concentration was 0.1998 mg/m^3 and it was observed in December at 10 out of 16 monitoring stations (Table A1). The most highly polluted regions were Peremogy prosp. (station 11), Besarabs'ka sq. (station 7), Lazo st. (station 4), and Lesi Ukrayinky blvd. (station 8). But even in these regions ambient level of PM did not exceed the maximum allowable concentration (Table 3.1).

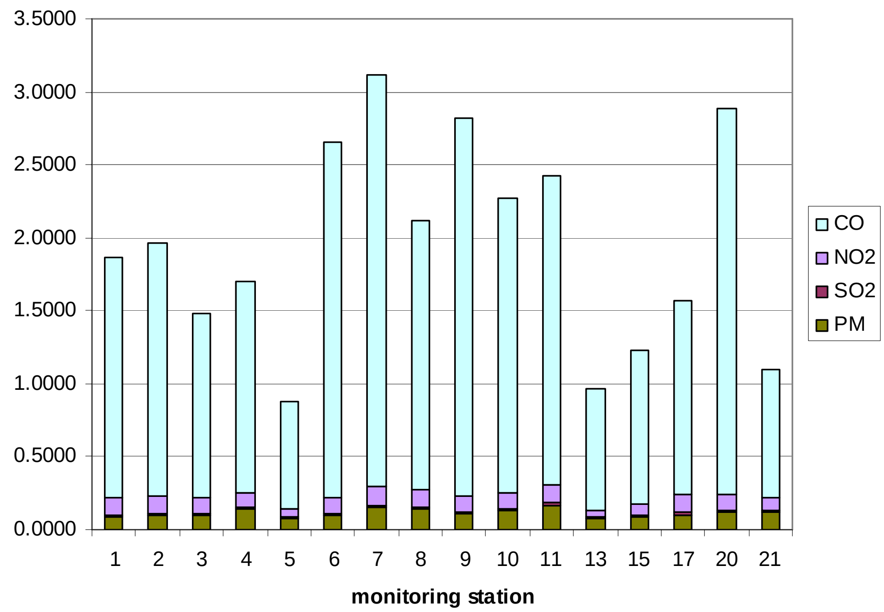
Ambient level of SO_2 also did not exceed maximum allowable concentration. This relatively low concentration of SO_2 could be explained by the slowdown in industrial production in the early period of transition of Ukraine. The annual average concentration of SO_2 ranged from 0.0103 (Popudrenko st., station 3) to 0.0165 mg/m^3 (Obolons'kyi prosp., station 17 and Moskovs'ka sq., station 20). Maximum observable concentration was 0.1298 mg/m^3, it was observed in June at Peremogy prosp (Table A1).

The annual average concentration of NO_2 ranged from 0.0501(Nauky prosp., station 5) to 0.1326 mg/m^3 (Besarabs'ka sq., station 7). The maximum concentration of 0.9798 mg/m^3 was observed in May at Besarabs'ka sq. At all monitoring stations there were cases exceeding maximum allowable concentration: at 13 out of 16 monitoring stations the frequency of such cases were more that 50%.

The annual average concentration for 2005 year of CO ranged from 0.7311 (Peremogy prosp., station 11) to 2.8168 mg/m^3 (Peremogy sq., station 6). The maximum concentration of 24.0000 mg/m^3 was observed in January at Besarabs'ka sq. (station 7) Cases exceeding maximum allowable concentration were observed at 13 out of 16 monitoring stations.

Overall, the most polluted regions are Peremogy sq., Besarabs'ka sq., Kaunas'ka st., and Lazo st. (Figure 3.2) with CO (1.7113 mg/m^3 annual average) contributing most to the air pollution of Kyiv.

Figure 3.2: Air pollution in Kyiv for 2005 year, annual average concentrations



Now we investigate the correlation between the variables on air pollution that is used for further analysis. Table 3.2 presents the correlation matrix between different pollutants.

Table 3.2: Matrix of correlation between different pollutants

	PM	SO_2	NO_2	CO
PM	1.0000			
SO_2	0.3782	1.0000		
NO_2	**0.5840**	0.4389	1.0000	
CO	**0.5329**	0.2971	**0.7078**	1.0000

The bold entries in correlation matrix show that variables are highly correlated. The correlation between carbon monoxide (CO) and nitrogen dioxide (NO_2) is high and constitute 0.7078, which reflect the fact that both air pollutants are generated mostly by vehicle emissions. The same explanation is valid for the high correlation between carbon monoxide (CO) and particulate matter (PM) (0.5329) and between particulate matter and nitrogen dioxide (0.5840). The association between PM and NO_2 is also accounted for the oxidation process of NO_2 in atmosphere that leads to creation of different particulate compounds. The results from the correlation analysis confirm the following observation: three out of four pollutants (PM, NO_2, and CO) can be considered as pollutants that come primary from mobile emission sources, while SO_2 – as pollutant that is generated mostly by stationary sources. Therefore, in order to overcome multicollinearity problem between different pollutants, we implement principal component analysis and search for the orthogonal principal components which explain most of the variation in the original data. We expect to find two principal components: one reflecting pollutants from mobile emission sources, and another reflecting pollutant from stationary sources.

The data on housing prices and structural characteristics consists of two data sets. First data set is the monthly pooled advertisements data for the years 2002-2005 collected by *Sioma* (2006). It contains information on advertisement rental prices for one-room apartments, main structural characteristics, and also the date of data collection. The data was collected using Aviso newspaper, which is published since January 1991 and represents the most popular advertisements newspaper. According to the information provided by the real estate agents Aviso newspaper covers approximately 85-90 percents of the secondary housing market.

The data contains 12665 observations (2002-2005 years). We take data only for the year 2005, which results in 4657 observations. Since there are only 16 monitoring stations in Kyiv, not all housing data will be used in the regression analysis; a sub-sample of houses is drawn from the area within 500 meter radius of the monitoring station. Thus, the final sample includes 2145 observations.

There are several limitations associated with this data (Sioma, 2006):

1. The set includes only one-room apartments and doesn't contain information on two- and three-room apartments.
2. The advertisement data is characterized by the lack of details. It means that not all characteristics of the apartments, which are actually present, are announced by the landlords. Landlords usually advertise such characteristics of apartments as address, number of rooms, total floor area, living area and area of the kitchen. Information about presence of the TV set, telephone, and refrigerator also appear very frequently. But all other characteristics, such as presence of the glassed balcony, presence of the armored door, presence of other appliances in the apartment, presence of the furniture, and state of the apartment, may appear in the advertisement or not even if they are actually present. Therefore, the data is constructed in such a way that if the characteristic is not advertised it is assumed to be unavailable.
3. The advertisement data is characterized by the subjectivity. Subjectivity refers to the selection and way of announcement of characteristics, which appear in the advertisement. It means that usually landlords choose those characteristics to advertise, which characterize the apartment from the best side. Subjectivity also can refer to the way of advertising. For example, landlords can overestimate the quality of some characteristics,

such as presence of the furniture and state of the apartment, or they can underestimate other characteristics. Therefore, this problem can result in errors in variables.

The neighborhood characteristics were constructed using the electronic business map of Kyiv. The variable **metro** represents the proximity to the metro station measured in meters, and the variable **park** shows the proximity to the green space labeled as **park** on the map. We also assign to each observation the characteristics of the environmental condition based on the location. Variables definitions are presented in the Table 3.3, descriptive statistics can be found in Table 3.4.

Table 3.3: Variable definitions, rental prices as dependent variable

Dependent Variable		
rent	Advertisement rental price of the apartment	(USD)
logrent	The natural logarithm of rent	
rent_perm	Advertisement rental price of the apartment per square meter	(USD/m^2)
Structural characteristics of the apartment		
area	Total floor area of the apartment	(m^2)
height	More-than-average height of the apartment	(1,0)*
balcony	Presence of the glassed balcony	(1,0)
door	Presence of the armored door	(1,0)
tel	Presence of the telephone	(1,0)
fridge	Presence of the fridge	(1,0)
tv	Presence of the TV set	(1,0)
appliances	Presence of other appliances in the apartment	(1,0)
furniture	Presence of the furniture	(1,0)
renovated	The state of the apartment is "after repair"	(1,0)
euroremont	The state of the apartment is "after Euro repair"	(1,0)
studio	The state of the apartment is "after specially designed Euro repair"	(1,0)
Location and characteristics of the neighborhood		
district	Location in one of the tenths districts	9 dummy variables for 10 districts: 1.Goloseeyivs'kyi, 2.Darnyts'kyi, 3.Desnians'kyi, 4.Dniprovs'kyi, 5.Obolons'kyi, 6.Pechers'kyi,

Table 3.3: Continued

		7.Podil's'kyi, 8.Shevchenkivs'kyi, 9.Solomyans'kyi, 10.Svyatoshyns'kyi
metro	Proximity to the metro station	(km)
park	Proximity to the green spaces	(km)
month	Time period	12 dummy variables

(1,0) indicates the dummy variable (1 – if the characteristic is available, 0 – if not)

Table 3.4: Descriptive statistics of variables, rental prices as dependent variable

Variable	Obs	Mean	Std. Dev.	Min	Max
rent	2145	311.6760	179.3945	80	2450
rent_perm	1691	9.0881	3.8304	2.0920	65.7143
area	1691	33.6434	7.6479	14	108
height	2145	0.0196	0.1386	0	1
balcony	2145	0.3287	0.4698	0	1
door	2145	0.3091	0.4622	0	1
tel	2145	0.7972	0.4022	0	1
tv	2145	0.4489	0.4975	0	1
appliances	2145	0.1678	0.3738	0	1
furniture	2145	0.8051	0.3962	0	1
renovated	2145	0.2890	0.4534	0	1
euroremont	2145	0.0904	0.2869	0	1
studio	2145	0.0252	0.1567	0	1
district	2145	5.3128	2.3475	1	10
metro	2145	950.1016	1160.0700	121	4673
park	2145	1176.7290	920.4845	179	2933
month	2145	6.6489	3.4420	1	12

The second data set was obtained from local real estate agency and is the advertisements data for the years 2004-2006. This data includes all information that appears in the Aviso newspaper. It contains information on advertised sales prices of the apartments (not rental prices), main structural characteristics, and also the date of data collection. Structural characteristics are the following: number of rooms, total floor area of the apartment, floor, presence of the telephone, presence of the hard wood floor, type of the bathroom, and wall material. This data set is also subject to the limitations (2) and (3) described above.

The data set contains 433230 observations covering the period 05.2004 – 03.2006. Since there are only 16 monitoring stations in Kyiv, again only a sub-sample of 6497 houses from the area within 500 meter radius of the monitoring station is used for the following analysis. One more problem associated with this data is that one apartment can appear several times in the advertisement newspaper Aviso. Therefore, number 6497 doesn't actually mean that the sample includes 6497 unique observations. Assuming that repeated observations are those that have the same address (street and building number), same number of rooms, and located on the same floor we can distinguish 1155 unique advertisements and 5342 repeated advertisements. From 5342 repeated observations we still can identify 1170 unique apartments. Taking only those with positive sales price (observations with zero price are indication of the errors and excluded from the sample) the sample results in 2081 observations. After considering data only for 2005 year, the final sample includes 1329 observations.

Variables definitions are presented in the Table 3.5 with variables describing neighborhood and environmental condition being constructed in the similar way as presented above. Descriptive statistics can be found in Table 3.6.

Table 3.5: Variables definitions, sales prices as dependent variable

Dependent Variable		
price	Advertisement sales price	(USD)
Structural characteristics of the apartment		
room	Number of rooms	4 dummies
area	Total floor area of the apartment	(m^2)
floor	Ground or upper floor	2 dummies for 3 cases: 1. ground floor, 2. upper floor, 3. interim floor
parquet	Presence of the hard wood floor	(1,0)
tel	Presence of the telephone	(1,0)
bath	Joint type of bathroom	(1,0)
wall	Wall material (brick/concrete)	(1,0)
Location and characteristics of the neighborhood		
the same as in Table 3.2		

Table 3.6: Descriptive statistics of variables, sales prices as dependent variable

Variable	Obs	Mean	Std. Dev.	Min	Max
price	1329	161316.7	201091.7000	15000	1830000
price_perm	1313	2012.3660	1584.9450	187.5000	31785.7100
room	1329	2.3664	1.0833	1	7
area	1329	70.4715	49.7422	0	540
floor	1201	0.2565	0.4369	0	1
parquet	580	0.9569	0.2033	0	1
tel	1329	0.5132	0.5000	0	1
bath	376	0.3378	0.4736	0	1
wall	1329	0.7705	0.4207	0	1
district	1329	6.0068	2.1813	1	10
metro	1329	898.5636	1160.0420	121	4673
park	1329	1269.0190	892.1145	179	2933
metrosq	1329	2152101	5231067	14641	2.18e+07
parksq	1329	2405678	2872735	32041	8602489
month	1329	15.5026	3.0545	9	20

Since variables **parquet** and **bath** have a large number of missing observations, they are not considered in the further analysis.

Chapter 4

METHODOLOGY

A. PRINCIPAL COMPONENT ANALYSIS

At the first stage of our research, we use principal component analysis (PCA) in order to reduce the dimension of the original data on air pollution and alleviate multicollinearity problem. We demonstrate PCA following the approach presented in Chartfield and Collins (1980). Let $X^T = [X_1, ..., X_p]$ is the initial set of p variables, which are correlated (in our case p data on different pollutants). Suppose that X^T has mean μ and covariance matrix $Var(X) = \Sigma$. It is assumed that variations and covariations in the given random variable X can be explained by the limited number of uncorrelated variables $Y_1, ..., Y_p$. Thus, the main aim of PCA is to derive these new orthogonal variables, principal components, and, thereafter, interpret them.

Each principal component Y_j can be represented by the linear combination of the original variables:

$$Y_j = a_{1j} X_1 + a_{2j} X_2 + ... + a_{pj} X_p = a_j^T \mathrm{X} \qquad (4.1)$$

with $a_j^T = [a_{1j}, ..., a_{pj}]$ being a vector of constants. To guarantee that the new set of variables $Y_1, ..., Y_p$ are indeed orthogonal, we require that $a_j^T a_j = \sum_{k=1}^{p} a_{kj}^2 = 1$.

The first principal component, Y_1, is calculated by maximizing the variance of Y_1:

$$\max \; Var(Y_1) = Var(a_1^T \mathrm{X}) = a_1^T \Sigma a_1 \qquad (4.2)$$

$$\text{s.t. } a_1^T a_1 = 1$$

The Lagrangian for this maximization problem is given by:

$$L(a_1) = a_1^T \Sigma a_1 - \lambda(a_1^T a_1 - 1) \qquad (4.3)$$

then, we obtain

$$\frac{\partial L}{\partial a_1} = 2\Sigma a_1 - 2\lambda a_1$$

Equalizing the last expression to **0,** we obtain

$$(\Sigma - \lambda \mathrm{I})a_1 = 0 \qquad (4.4)$$

then we find the solution by setting

$$|\Sigma - \lambda \mathrm{I}| = 0 \qquad (4.5)$$

if we do not allow for a_1 being the null vector. Looking more closely at the expression (4.5), we find that this is exactly the definition of λ being the eigenvalue of Σ. Therefore, in order to maximize the variance of Y_1:

$$Var(Y) = Var(a_1^T \mathrm{X}) = a_1^T \Sigma a_1 = a_1^T \lambda \mathrm{I} a_1 = \lambda$$

we set λ equal to the largest possible eigenvalue λ_1 of Σ and a_1 can be found as an eigenvector corresponding to λ_1.

Using the same procedure we can find the second principal component Y_2, but in this case we impose additional constraint, namely that Y_2 is orthogonal to Y_1:

$$Cov(Y_2, Y_1) = Cov(a_2^T X, a_1^T X) = E\left[a_2^T (X - \mu)(X - \mu)^T a_1\right] = a_2^T \Sigma a_1 \qquad (4.6)$$

To ensure that Y_2 is uncorrelated with Y_1 we set $Cov(Y_2, Y_1) = 0$ or $a_2^T a_1 = 0$ (using (4.4)). Thus, the maximization problem becomes:

$$\max\ Var(Y_2) = Var(a_2^T \mathrm{X}) = a_2^T \Sigma a_2$$

$$\text{s.t. } a_2^T a_2 = 1$$

$$a_2^T a_1 = 0$$

The Lagrangian function in this case is given by:

$$L(a_2) = a_2^T \Sigma a_2 - \lambda(a_2^T a_2 - 1) - \delta a_2^T a_1 ,$$

and

$$\frac{\partial L}{\partial a_2} = 2(\Sigma - \lambda \mathrm{I})a_2 - \delta a_1 = 0 \qquad (4.7)$$

Premultiplying the last expression by a_1^T and using the result $a_1^T a_2 = 0$ we receive $2a_1^T \Sigma a_2 - \delta = 0$. Since $a_1^T \Sigma a_2$ equals to zero by (4.6), δ should also be zero at the stationary point. Thus, equation (4.7) now can be written as:

$$(\Sigma - \lambda \mathrm{I})a_2 = 0$$

The obtained equation is similar to the one that we saw above (4.4). Therefore, in order to maximize the variance of Y_2:

$$Var(Y_2) = Var(a_2^T \mathrm{X}) = a_2^T \Sigma a_2 = a_2^T \lambda \mathrm{I} a_2 = \lambda$$

we set λ equal to the second largest eigenvalue λ_2 of Σ and a_2 can be found as an eigenvector corresponding to λ_2. Following the same procedure, we find all other principal components as eigenvectors corresponding to the respective eigenvalues.

Denoting the $(p \times p)$ matrix of eigenvectors by A, $A = \lfloor a_1, ..., a_p \rfloor$, vector of principal components by Y we can rewrite Y as

$$Y = A^T X$$

and the variance of Y, $Var(Y) = \Lambda$, is given by $\Lambda = A^T \Sigma A$, where

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ ... & & & \\ 0 & ... & ... & \lambda_p \end{pmatrix}$$

Thus, the sum of the variances becomes:

$$\sum_{i=1}^{p} Var(Y_i) = \sum_{i=1}^{p} \lambda_i = trace(\Lambda)$$

On the other hand,

$$trace(\Lambda) = trace(A^T \Sigma A) = trace(\Sigma A A^T) = trace(\Sigma) = \sum_{i=1}^{p} Var(X_i).$$

Therefore, the sum of the variances of the original variables equal to the sum of the variances of the principal components, which allows to interpret $\frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j}$ as the proportion of the total variation of the original data explained by the i th principal component.

B. REGRESSION ANALYSIS

At the second stage of our research, we will estimate six different regressions:

(i) linear Box-Cox model with transformed dependent and independent variables using different parameters

$$p_i(\lambda) = \alpha_0 + \sum_{j=1}^{k} \alpha_j x_{ij}(\mu) + \sum_{l=k+1}^{n-1} \alpha_l x_{il} + \alpha_n x_{in} + \varepsilon_i \ ,$$

where $p_i(\lambda) = \frac{p_i^\lambda - 1}{\lambda}$ for $\lambda \neq 0$. If λ equals to zero, the transformation becomes $\lim_{\lambda \to 0} \frac{p_i^\lambda - 1}{\lambda} = \lim_{\lambda \to 0} \frac{d(p_i^\lambda - 1)/d\lambda}{1} = \lim_{\lambda \to 0} p_i^\lambda \times \ln p_i = \ln p_i$, where the L'Hopial rule was used (variables x_{ij} are transformed in the same way).

p_i – rental (or sales) price for the i^{th} apartment;

$x_{ij}, j = 1,...,k$ – structural characteristics of the apartment and characteristics of the neighborhood that can be transformed (**area, metro, park**);

α_j – parameters associated with x_{ij};

$x_{il}, l = k+1,...,n-1$– other structural characteristics of the apartment and characteristics of the neighborhood that can not be transformed;

α_l – parameters associated with x_{il};

x_{in} – principal component that represents pollution;

α_n – parameter associated with x_{in} (this is the coefficient of our interest);

ε_i – independently and identically distributed error term.

(ii) linear Box-Cox model with transformed dependent and independent variables using the same parameter

$$p_i(\lambda) = \alpha_0 + \sum_{j=1}^{k} \alpha_j x_{ij}(\lambda) + \sum_{l=k+1}^{n-1} \alpha_l x_{il} + \alpha_n x_{in} + \varepsilon_i \ ;$$

(iii) linear Box-Cox model with transformed independent variables (λ = 1)

$$p_i = \alpha_0 + \sum_{j=1}^{k} \alpha_j x_{ij}(\mu) + \sum_{l=k+1}^{n-1} \alpha_l x_{il} + \alpha_n x_{in} + \varepsilon_i \ ;$$

(iv) linear Box-Cox model with transformed dependent variable(μ = 1)

$$p_i(\lambda) = \alpha_0 + \sum_{j=1}^{n-1} \alpha_j x_{ij} + \alpha_n x_{in} + \varepsilon_i \ ;$$

(v) semi-log model (λ = 0, μ = 1)

$$\ln p_i = \alpha_0 + \sum_{j=1}^{n-1} \alpha_j x_{ij} + \alpha_n x_{in} + \varepsilon_i \ ;$$

(vi) linear model (λ = 1, μ = 1)

$$p_i = \alpha_0 + \sum_{j=1}^{n-1} \alpha_j x_{ij} + \alpha_n x_{in} + \varepsilon_i \ .$$

First four models, Box-Cox models, will be estimated using the Maximum Likelihood method. The estimation procedure for the second model can be found in Appendix C. Other models are estimated in a similar way.

Estimated hedonic price equations can be used to find the welfare measures. For the linear Box-Cox model *MWTP* for the characteristic of our interest x_n can be calculated in the following way

$$\frac{\partial p_i}{\partial x_{in}} = p_i^{1-\lambda} \alpha_n x_{in}^{\lambda-1}.$$

WTP for the improvement in air quality is the following

$$WTP = f(\vec{x}_i^{\,*}) - f(\vec{x}_i)$$

using $f(\vec{x}^*) = \overline{p}$ and $\frac{p^\lambda - 1}{\lambda} = \vec{x}(\lambda)\vec{\alpha}$ we can express $f(\vec{x}^*) = (\overline{p}^\lambda + (\lambda x_j^*(\lambda) - x_j(\lambda))\alpha_j)^{1/\lambda}$.For other forms of the Box-Cox model, the solutions are less trivial, but are easily calculated on the computer.

Chapter 5

ESTIMAION RESULTS

A. PRINCIPAL COMPONENT ANALYSIS

As the first stage of our analysis, we implement principal component analysis (PCA) using data on air pollution in Kyiv for 2005 year. While 4 pollutants (PM, SO_2, NO_2, and CO) considered in the current analysis represent different dimensions of the air quality, they are not independent of each other. The reason of such independence could be a common source of emission. Therefore, we use PCA in order to reduce the dimension of our original data and construct measures of air quality that are uncorrelated.

Following the procedure described in Chapter 4, we calculate eigenvalues and eigenvectors for the correlation matrix of original data. The results are presented in Tables 5.1.

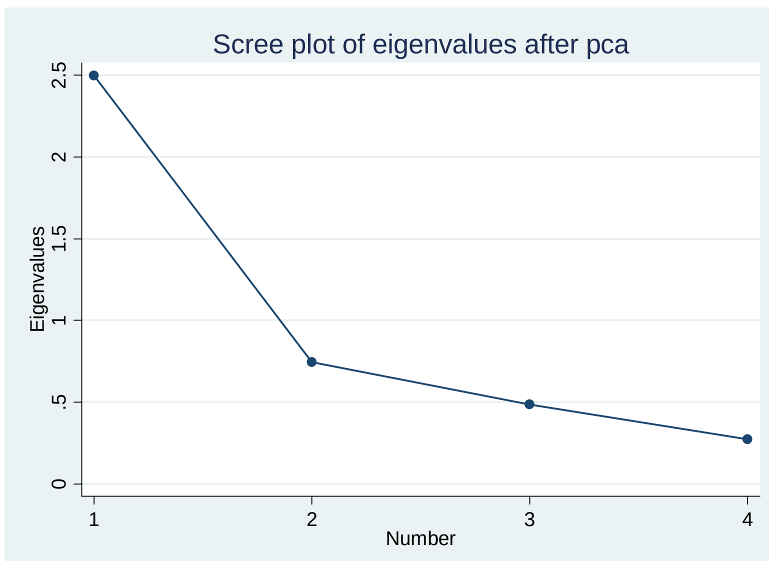
Table 5.1: Eigenvalues and proportion of variance explained by principal components

Principal component	Eigenvalue	Difference	Proportion	Cumulative
Y_1	2.4970	1.7519	0.6242	0.6242
Y_2	0.7451	0.2594	0.1863	0.8105
Y_3	0.4857	0.2135	0.1214	0.9319
Y_4	0.2722	.	0.0681	1.0000

The first principle component (Y_1) corresponds to the largest eigenvalue $\lambda_1 =$ 2.4970 and explains 62.42% of the total variance. The second principal component corresponds to the second largest eigenvalue $\lambda_2 =$ 0.7451 and explain 18.63% of the total variance. Therefore, Y_1 and Y_2 explain 81.05% of the variation.

Plotting eigenvalues (Figure 5.1) allows us to see graphically the above results. Nevertheless smooth decrease of eigenvalues starts to level off after the first eigenvalue, which is an indicator for retaining only one principal component for subsequent analysis; we are going to retain two of them in order to be consistent with theoretical findings.

Figure 5.1: Scree plot of eigenvalues, PCA



Corresponding eigenvectors are given in Table 5.2.

Table 5.2: Principal components, eigenvectors

Variable	Y_1	Y_2	Y_3	Y_4
PM	0.5050	-0.0930	0.8554	0.0670
SO_2	0.3959	0.8897	-0.1503	0.1705
NO_2	0.5606	-0.1521	-0.2879	-0.7614
CO	0.5234	-0.4202	-0.4034	0.6219

They are essentially the weights of the original variables in the principal components. For example:

Y_1 = 0.5050 PM + 0.3959 SO_2 + 0.5606 NO_2+ 0.5234 CO;

Y_2 = -0.0930 PM + 0.8897 SO_2 - 0.1521 NO_2 - 0.4202 CO.

Therefore, first principal component is actually the sum of PM, NO_2, and CO; while second principal component represents mostly SO_2 pollutant. To receive better interpretation of the principal components we look at correlation between principal components and original variables (Table 5.3).

Table 5.3: Correlation between principal components and original variables

Variable	Y_1	Y_2	Y_3	Y_4
PM	0.7981	-0.0803	0.5962	0.0350
SO_2	0.6255	0.7680	-0.1047	0.0890
NO_2	0.8858	-0.1313	-0.2006	-0.3973
CO	0.8271	-0.3627	-0.2811	0.3245

We have already interpreted first principal component as the sum of PM, NO_2, and CO; this is also reflected in the Table 5.3. The correlation between Y_1 and PM is 0.7981, between Y_1 and NO_2 it is 0.8858, between Y_1 and CO it is 0.8271. Second principal component represent mostly SO_2 pollutant, which is also represented in Table 5.3: the correlation between Y_2 and SO_2 is 0.7680, while correlation between Y_2 and pollutants from mobile emission sources is negative. Thus, we will use two principal components, Y_1 that represent pollutants from mobile emission sources, and Y_2 than reflects pollutant from stationary emission sources, for our further analysis (Stata coding for Y_1 and Y_2 is **pc(mobile)** and **pc(stationary)** respectively).

B. REGRESSION ANALYSIS

ANALYSIS OF HOUSING RENTAL PRICES

We start regression analysis with the simplest linear model using data on rental prices for 2005 year. Then we test linear model versus semi-log model using test proposed by MacKinnon, White, and Davidson. We assume that

H_0: linear model: $p_i = \alpha_0 + \sum_{j=1}^{n-1} \alpha_j x_{ij} + \alpha_n x_n + \varepsilon_i$;

H_1: semi-log model: $\ln p_i = \alpha_0 + \sum_{j=1}^{n-1} \alpha_j x_{ij} + \alpha_n x_n + \varepsilon_i$.

First we estimate linear model (Table 5.4) and obtain fitted values $\hat{p}_i$ (we call them $\hat{p}_i f$), then we estimate semi-log model (Table 5.4) and obtain fitted values, which we call $\ln \hat{p}_i$. Introducing the new parameter $\ln \hat{p}_i f - \ln \hat{p}_i$ into the linear form model, we check whether it is significantly different from zero using *t* test. The coefficient on new parameter $\ln \hat{p}_i f - \ln \hat{p}_i$ is -5.2740 with t-statistic being -5.4600. So, we reject the null hypothesis of linear functional form in favour of semi-log model.

Table 5.4: Estimation results, linear and semi-log regression models, rental prices per square meter as dependent variable

	Linear model	**Semi-log model**	**Semi-log model with some variables squared**	**Semi-log model without influential observations**
pc(mobile)	-0.3324*** (-0.1188)	-0.0493*** (-0.0089)	-0.0814*** (-0.0100)	-0.0731*** (-0.0089)
pc(stationary)	-1.3011*** (-0.1414)	-0.1094*** (-0.0106)	-0.1607*** (-0.0122)	-0.1612*** (-0.0113)

Table 5.4: Continued

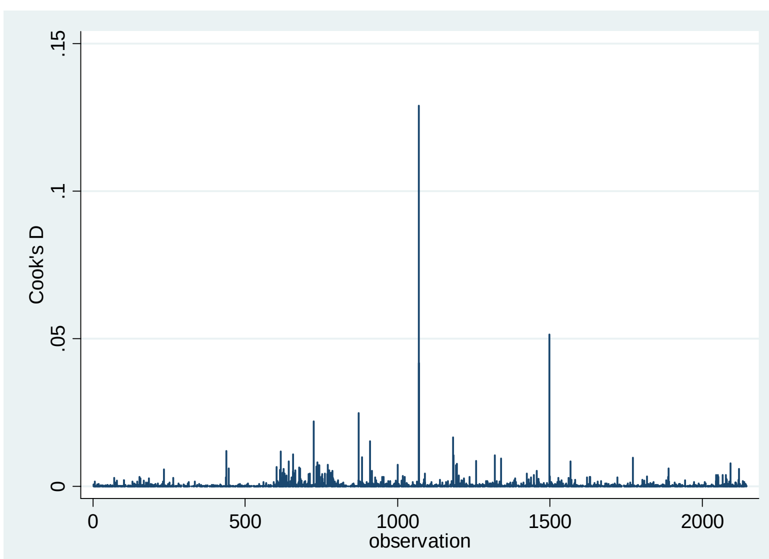
area	-0.0845*** (-0.0092)	-0.0116*** (-0.0007)	-0.0196*** (-0.0023)	-0.0169*** (-0.0023)
areasq	–	–	0.0001*** (2.31e-05)	3.69e-05 (2.38e-05)
height	0.5576 (-0.4694)	0.0581* (-0.0352)	0.0546 (-0.0345)	0.0614* (-0.0328)
balcony	-0.0389 (-0.1616)	-0.0009 (-0.0121)	-0.0025 (-0.0119)	-0.0038 (-0.0113)
door	-0.1372 (-0.1488)	-0.0031 (-0.0112)	0.0003 (-0.0109)	0.0049 (-0.0104)
tel	-0.1197 (-0.1759)	-0.0037 (-0.0132)	-0.0043 (-0.0129)	-0.0054 (-0.0123)
tv	0.4309*** (-0.1454)	0.0385*** (-0.0109)	0.0379*** (-0.0107)	0.0328*** (-0.0102)
appliances	1.2751*** (-0.2001)	0.1408*** (-0.0150)	0.1375*** (-0.0147)	0.1378*** (-0.014)
furnture	-0.0097 (-0.1823)	0.0012 (-0.0137)	0.0106 (-0.0134)	0.0062 (-0.0128)
renovated	0.5690*** (-0.1496)	0.0691*** (-0.0112)	0.0666*** (-0.0110)	0.0709*** (-0.0105)
euroremont	3.8140*** (-0.2905)	0.3362*** (-0.0218)	0.3385*** (-0.0215)	0.3525*** (-0.0206)
studio	2.7445*** (-0.5481)	0.1900*** (-0.0411)	0.1793*** (-0.0404)	0.1175*** (-0.0393)
metro	-0.4053*** (-0.1138)	-0.0545*** (-0.0085)	-0.0676*** (-0.0222)	-0.0824*** (-0.0088)
metrosq	–	–	-0.0051 (-0.0052)	–
park	0.7427*** (-0.1860)	0.0860*** (-0.0140)	0.5249*** (-0.0580)	0.5190*** (-0.0537)
parksq	–	–	-0.1535*** (-0.0202)	-0.1534*** (-0.0183)
Constant	10.7285*** (-0.6518)	2.3862*** (-0.0489)	2.5690*** (-0.0682)	2.5448*** (-0.0660)
Observations	1691	1691	1691	1688
R-squared	0.53	0.66	0.68	0.7
F-statistics for district dummies	40.21	78.45	44.72	51.14
F-statistics for month dummies	31.99	71.81	72.71	78.09

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

We check whether there is a linear relationship between rental prices (to be precise rental prices per square meter) and area, and also between rental prices and variables that reflect proximity to the metro station and green spaces. Including area squared (**areasq**), proximity to the metro station squared (**metrosq**), and proximity to the green spaces squared (**parksq**) into the semi-log model (Table 5.4), we find that coefficients on **areasq** (0.0001) and **parksq** (-0.1535) are statistically significant at 5% level (Table 5.4), while coefficient on **metrosq** (-0.0051) is not statistically significant at 5% level. LR test confirms our findings that area squared and proximity to the green spaces squared should be included into regression in square form (LR statistics = 75.6700).

We check the semi-log model for misspecification and robustness by finding influential observations first. To identify influential cases we use Cook's Distance. Figure 5.2 shows that observations 1069, 1070, and 1498 have large Cook's distances.

Figure 5.2: Cook's Distance from the semi-log model, measure of overall influence



To find the reasons for influence we look at outliers (Figure 5.3) and leverage (Figure 5.4). Large studentized residuals for observations 1069, 1070, and 1498 indicate that these observations are outliers.

Figure 5.3: Studentized residuals for detecting outliers from the semi-log model

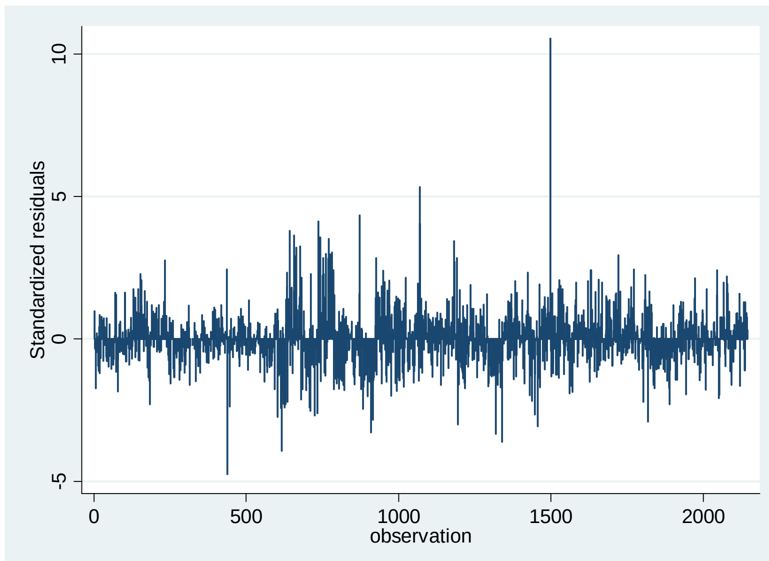
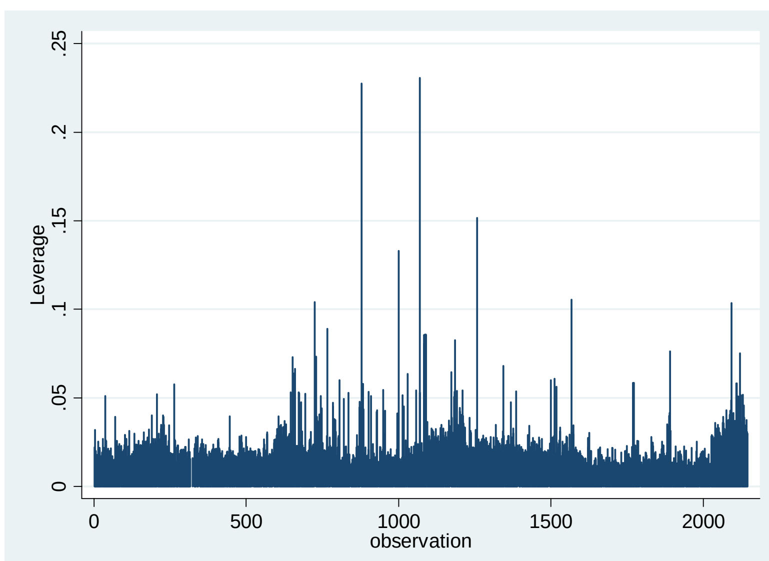


Figure 5.4: Leverage from the semi-log model



Observation 1070 has also very large leverage (Figure 5.4). Therefore, we exclude observations 1069, 1070, and 1498 from our sample. After exluding influential observations (Table 5.4) we check the semi-log model for misspecification and robustness finding that the model suffers from non-normality and heteroskedasticity. The probability that residuals from semi-log model normally distributed is less than 0.0010 based on Skewness/Kurtosis test for normality. We also reject the null hypothesis that residuals are homoskedastic (i.e. they have constant variance) on the basis of the Breusch-Pagan/Cook-Weisberg test. $\chi^2(1)$-statistics equals to 110.19. One solution to this problem is the Box-Cox transformation of dependent and, possibly, independent variables.

We use the Atkinson Score test to examine whether we need to transform dependent variable (rental prices per square meter). Suppose that the true model is

$$p_i(\lambda) = \alpha_0 + \sum_{j=1}^{n-1} \alpha_j x_{ij} + \alpha_n x_n + \varepsilon_i \ .$$

Using the first-order Taylor series approximation around $\lambda = 1$:

$$p(\lambda) = p(1) + (\lambda - 1)\frac{dp(\lambda)}{d\lambda}\Big|_{\lambda=1} ,$$

where $\frac{dp(\lambda)}{d\lambda}\Big|_{\lambda=1} = a + \log \tilde{p} + 1$ with $a_i = p_i(\log\left(\frac{p_i}{\tilde{p}}\right) - 1)$, and $\tilde{p}$ being the geometric mean of p. Therefore, including additional variable a_i into the linear model:

$$p_i = \alpha_0 + \sum_{j=1}^{n-1} \alpha_j x_{ij} + \alpha_n x_n + \gamma a_i + \varepsilon_i$$

and test whether the coefficient on a_i is significantly different from zero, we can infer whether we need the transformation of dependent variable. We estimate that coefficient on a_i equals 1.2401 and is statistically significant from zero at 5% level (Table 5.5). Thus, we need the transformation of dependent variable.

We also check whether we need to transform the continuous independent variables (**area, metro, park**). We include into the linear model additional variables:

area_at = area $*$ log(area)

metro_at = metro $*$ log(metro)

park_at = park $*$ log(park)

and test whether the estimated coefficients on this variables are significant (Table 5.5). The coefficient on **area_at** equals to 0.04745 (with t-statistics equal to 2.8100), on **metro_at** it is -0.1570 (with t-statistics equals to -1.1000), on **park_at** it is -2.2174 (with t-statistics equals to -7.0100). Coefficients on **area_at** and **park_at** are statistically significant from zero at 5% level. Therefore, we conclude that we need to transform both the independent and dependent variables. Model which assumes only the transformation of the dependent variable (lhs), and model which assumes the transformation of the independent variables (rhs) do not perform well enough in explaining our data (Table 5.6).

Table 5.5: Estimation results, linear model, rental prices per square meter as dependent variable, testing for the need of variables transformation

	Linear model Test for the need of dependent variable transformation	**Linear model Test for the need of independent variables transformation**
pc(mobile)	-0.4595*** (-0.0626)	-0.6679*** (-0.0696)
pc(stationary)	-0.5971*** (-0.0763)	-0.9912*** (-0.0916)
area	-0.0991*** (-0.0050)	-0.3319*** (-0.0809)

Table 5.5: Continued

height	0.3899 (-0.2468)	0.3441 (-0.2419)
balcony	-0.0120 (-0.0850)	-0.0180 (-0.0834)
door	0.0488 (-0.0783)	0.0743 (-0.0768)
tel	-0.0283 (-0.0924)	-0.0293 (-0.0906)
tv	0.2319*** (-0.0764)	0.2320*** (-0.0750)
appliances	1.1527*** (-0.1052)	1.1236*** (-0.1033)
furniture	0.0009 (-0.0958)	0.0640 (-0.0942)
renovated	0.5865*** (-0.0787)	0.5692*** (-0.0772)
euroremont	2.1277*** (-0.1589)	2.1427*** (-0.1562)
studio	0.5464* (-0.2965)	0.4992* (-0.2910)
metro	-0.4836*** (-0.0599)	-0.5253** (-0.2249)
park	0.6953*** (-0.0977)	3.3666*** (-0.3826)
a_i	1.2401*** (-0.0254)	1.2180*** (-0.0250)
area_at	–	0.0475*** (-0.0169)
metro_at	–	-0.1570 (-0.1427)
park_at	–	-2.2174*** (-0.3161)
Constant	20.8649*** (-0.3950)	22.1096*** (-0.80710)
Observations	1688	1688
R-squared	0.85	0.85
Standard errors in parentheses		
significant at 10%; ** significant at 5%; *** significant at 1% All models include dummies for month and district		

Table 5.6 reports the estimation results for the Box-Cox model. We focus on the last column, which presents estimates for the Box-Cox model with transformed

dependent and independent variables. The parameter of transformation (**lambda**) is significantly different from zero at 5% level. This is an indication that flexible functional form of hedonic price equation performs better in terms of goodness-of-fit.

Table 5.6: Estimation results, Box-Cox models, rental prices per square meter as dependent variable

	Box-Cox model lhs[1]	**Box-Cox model lhs+rhs**
pc(mobile)	-0.02826*** (36.17500)	-0.02224*** (22.51400)
pc(stationary)	-0.06096*** (115.60000)	-0.07172*** (100.85600)
area	-0.00758*** (362.94400)	-0.69215*** (347.76700)
height	0.03844** (4.34500)	0.03796* (3.18800)
balcony	-0.00123 (0.03800)	0.00018 (0.00100)
door	0.00244 (0.17400)	0.00368 (0.29700)
tel	-0.00035 (0.00200)	0.00188 (0.05600)
tv	0.01956*** (11.71700)	0.0218*** (10.93200)
appliances	0.07851*** (97.08400)	0.09277*** (101.78800)
furniture	-0.00058 (0.00700)	0.00229 (0.07700)
renovated	0.0421*** (50.59500)	0.04862*** (50.54900)
euroremont	0.19059*** (253.10100)	0.22226*** (256.26200)
studio	0.06145*** (7.72300)	0.05594** (4.82100)
metro	-0.03105*** (47.44300)	-0.03306*** 35.611
park	0.04827*** (43.19700)	0.06211*** (77.55600)

Table 5.6: Continued

Constant	1.80006	3.41099
lambda	-0.26708*** (0.03716)	-0.20248*** (0.03505)
mu	–	-0.20248*** (0.03505)
Log likelihood	-3161.54170	-3170.22890

Observations 1688
LR statistic in parentheses
For transformation coefficients standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
All models include dummies for month and district
[1]lhs=transformation of the left-hand side (dependent variable)
rhs=transformation of the right-hand side (independent variables)
Estimation of the Box-Cox model with transformed dependent and independent variables using different parameters did not converge.

Regarding the estimated parameters, total floor area of the apartment is negatively associated with the rental price per square meter of the apartment; while more-than average height of the apartment, TV set and other appliances, state of the apartment (after repair, after Euro repair, after specially designed Euro repair) are positively related to the rental price per square meter of the apartment. Other structural characteristics of the apartment: presence of the glassed balcony, armored door, presence of the telephone, and furniture are not significantly associated with the rental price. The insignificance of the presence of the telephone and furniture can be explained by the following reasoning. Since both these variables have large average values (Table 3.4), and missing data on these variables is coded as the absence of the above mentioned features, we can assume that the true proportion of the apartments with telephone and furniture is higher. Thus, these variables would have lower variation and their effect cannot be estimated precisely.

As to the location and characteristics of the neighbourhood we can conclude that remoteness from the metro station is negatively associated with the rental prices per square meter, which means that the farther away the apartment is located from the metro station, the lower will be the rental price. On the other hand, the effect of the green spaces has the opposite sign. This positive relationship is observed in all models and can be explained by the presence of omitted variables. If omitted variables are positively associated with the proximity to the green spaces, the coefficient under investigation could be biased upward. Location in Darnyts'kyi (district 2), Desnians'kyi (district 3), Dniprovs'kyi (district 4), and Svyatoshyns'kyi (district10) districts is negatively related to the rental price comparing with the location in Obolons'kyi district (distict 5), which is the reference one. On the other hand, location in Pechers'kyi (district 6) and Shevchenkivs'kyi (district 8) districts is positively related to the rental price (coefficients on district dummies are presented in the Table C1). These findings are consistent with statistic on average rental prices per square meter provided in Table 5.6.

Table 5.6: Average rental price per square meter ($) by districts of Kyiv, 2005 year

Goloseeyivs'kyi (district 1)	8.0204	Pechers'kyi (district 6)	12.5581
Darnyts'kyi (district 2)	7.1750	Podils'kyi (district 7)	9.4113
Desnians'kyi (district 3)	7.1404	Shevchenkivs'kyi (district 8)	10.6690
Dniprovs'kyi (district 4)	8.2205	Solomyans'kyi (district 9)	8.0722
Obolons'kyi (reference district)	8.0510	Svyatoshyns'kyi (district 10)	6.9993

The coefficients of our interest are on **pc(mobile)** and **pc(stationary),** which show the affect of the pollution on the rental prices per square meter. These coefficients have the right sign, which means that if the apartment is located in a more polluted area, the price will be lower. Estimated hedonic price equation can be used to find the welfare measures. For the linear Box-Cox model mariginal implicit price for the characteristic of our interest x_n can be calculated in the following way

$$\frac{\partial p}{\partial x_n} = p^{1-\lambda}\alpha_n x_n^{\lambda-1}.$$

Marginal implicit price can also be interpreted as marginal willingness to pay (MWTP) for an improvement in air quality. Assuming that p = 9.03395 (average rental price per square meter) we find that

Marginal value for **pc(mobile)** = $9.03395^{1.20248}$ (-0.02224) = -\$0.31373 or -3.47% of rental price per square meter (minus sign here is an indication that coefficient on **pc(mobile)** goes with minus sign);

Marginal value for **pc(stationary)** = $9.03395^{1.20248}$ (-0.07172) = -\$1.01173 or -11.20% of rental price per square meter (Table 5.7).

Table 5.7: Marginal implicit prices, rental prices

		Linear model	**Semi-log model**	**Box-Cox lhs**	**Box-Cox lhs+rhs**
Estimated coefficient	**pc(mobile)**	-0.33240***	-0.0731***	-0.02826***	-0.02224***
	pc(stationary)	-1.30110***	-0.1612***	-0.06096***	-0.07172***
Lambda		–	–	-0.26708***	-0.20248***
Sample mean		9.03395	9.03395	9.03395	9.03395
Marginal value[1]	**pc(mobile)**	-0.33240***	-0.66038***	-0.45957***	-0.31373***
	pc(stationary)	-1.30110***	-1.45627***	-0.99133***	-1.01173***
% of housing price	**pc(mobile)**	-3.68***	-7.31***	-5.09***	-3.47***
	pc(stationary)	-14.40***	-16.07***	-10.97***	-11.20***

[1]MWTP = $\frac{\partial p}{\partial x_n} = p^{1-\lambda}\alpha_n x_n^{\lambda-1}$,

Here we define marginal change to be 1 standard deviation (std) change in principal component. Since 1 std change in principal component that represents pollution from mobile sources, **pc(mobile)**, corresponds to the 0.5 std change in PM (0.01375 mg/m^3), 1 std increase is approximately a 12.2% reduction in mean PM concentrations. Therefore, the obtained result suggests that a household in the apartment with mean rental price (\$303.37) would like to pay approximately \$0.31 for each square meter of the apartment (or 3.47% of the housing price) to avoid a 12.2% increase in mean PM. In the similar way we find that 1 std increase in principal component that represents pollution from stationary sources, **pc(stationary)**, corresponds to the 0.9 std increase in SO_2 (0.0018 mg/m^3) or approximately 13.5% reduction in mean SO_2 concentrations. Thus, the obtained marginal value suggests that a household in the apartment with mean rental price (\$303.37) would like to pay approximately \$1.01 for each square meter of the apartment (or 11.20% of the housing price) to avoid a 13.5% increase in mean SO_2. Therefore, based on the analysis of rental housing prices, we conclude that people in Kyiv on average would like to pay more for the reduction of air pollution that comes from stationary sources comparing to the pollution that comes from mobile emission sources.

But we should very careful interpret the estimated marginal effects. Pollution from mobile sources could be positively associated with some unobservable factors such as good road interchanges and, as a result, better access to the jobs and social amenities. Since we do not control for these characteristics and we expect positive relationship between them and sales prices, the obtained coefficient on **pc(mobile)** could be overestimated and could be interpreted as the upper bound of the true coefficient on **pc(mobile)**. Pollution from stationary sources, on the other hand, could be associated with such factors as presence of the factory stacks, which are negatively associated with the sales prices. Therefore, coefficient on **pc(stationary)** could be interpreted as the lower bound of the true coefficient.

For the sake of comparison Table 5.7 also reports marginal values for other models. In all models ccoefficients on **pc(mobile)** and **pc(stationary)** go with the minus sign supporting the idea that households would like to pay a positive amount in order to avoid increase in air pollution. It is also interesting to note that marginal values from the simple linear model are quite close to the marginal values from the Box-Cox model, which supports findings of *Cropper* et al. (1988) that simple models provide more accurate estimates of marginal implicit prices comparing to others models if there is a problem of omitted variables.

ANALYSIS OF HOUSING SALES PRICES

We analyse housing sales prices using the same procedure as for the analysis of the housing rental prices. Table 5.9 reports the estimates for the Box-Cox model with transformed dependent variable. Coefficient of transformation (**lambda**) is significantly different from zero at 5% level suggesting that this model performs well enough in terms of goodness-of-fit (complete analysis of the housing sales prices can be found in Appendix D).

Table 5.8: Estimation results, Box-Cox model with transformed dependent variable, sales prices per square meter as dependent variable

pc(mobile)	0.00898*** (56.84000)	**pc(stationary)**	-0.01764*** (43.17400)
room1	0.01475*** (12.08100)	**district1**	-0.01567* (5.39500)
room2	0.01055*** (7.79700)	**district2**	-0.07658*** (72.13300)
room3	0.00408 (1.47900)	**district3**	-0.04564*** (41.47400)
area	0.00015*** (20.48200)	**district4**	-0.06621*** (187.24700)
floor1	-0.00887*** (11.67400)	**district5**	-0.00844 (1.33900)
floor2	-0.01371*** (33.48600)	**district7**	-0.02433*** (29.02300)
tel	-0.00613*** (11.93800)	**district8**	0.00061 (0.04900)
wall	0.00124 (0.21300)	**district9**	-0.03473*** (56.76100)

Table 5.8: Continued

Metro	0.00004*** (12.35500)	**district10**	-0.03306*** (37.59700)
park	0.00461* (4.90200)		
Constant	3.00529		
lambda	-0.29649*** (0.03757)		
Log likelihood	-8364.61330		

Observations 1125
LR statistic in parentheses
For transformation coefficient standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
The model includes dummies for month

One- and two-room apartments have higher sales price per square meter comparing with the four-room apartments (reference dummy). Positive association is also found for the total floor area of the apartment, while location of the apartment at the ground or upper floor is negatively related to the sales price per square meter comparing with the location at the interim floor. Negative association is also found for the presence of the telephone in the apartment.

As to the location and characteristics of the neighbourhood we find that proximity to the metro station is positively related to the sales prices per square meter. This positive relationship is observed in all specifications (Appendix D) and can be explained by the presence of omitted variables. If omitted variables are positively associated with the proximity to the metro station, the coefficient under investigation could be biased upward. Location in Darnyts'kyi (district 2), Desnians'kyi (district 3), Dniprovs'kyi (district 4), Podils'kyi (district 7), Solomyans'kyi (district 9), and Svyatoshyns'kyi (district10) districts is negatively related to the rental price comparing with the location in Pechers'kyi district (distict 6), which is the reference one. These findings are consistent with statistic on average sales prices per square meter provided in Table 5.9.

Table 5.9: Average sales price per square meter in $ by districts of Kyiv, 2005 year

Goloseeyivs'kyi (district 1)	1371.8340	Pechers'kyi (reference district)	2776.0990
Darnyts'kyi (district 2)	1062.4830	Podils'kyi (district 7)	1610.3290
Desnians'kyi (district 3)	1311.8500	Shevchenkivs'kyi (district 8)	2287.9110
Dniprovs'kyi (district 4)	1103.7730	Solomyans'kyi (district 9)	1234.2470
Obolons'kyi (district 5)	1376.8310	Svyatoshyns'kyi (district 10)	1252.5110

The coefficients of our interest are on **pc(mobile)** and **pc(stationary),** which show the affect of the pollution on the sales prices per square meter. The coefficient on **pc(mobile)** goes with the positive sign and is significant at 5% level, while **pc(stationary)** is negatively and significantly associated with the sales prices. This difference can also be explained using the argument presented above. Marginal value of pollutants from mobile emission sources could be interpreted as the upper bound of the true marginal value, while marginal value of pollutants from stationary emission sources could be seen as the lower bound of the true effect due to the presence of some unobservable factors.

Now we calculate marginal implicit prices for the characteristics of our interest. Assuming that p = 1909.67800 (average sales price per square meter) we find that

Marginal value for **pc(mobile)** = $1909.67800^{1.29649}$ (0.00898) = $161.06766 or 8.43% of sales price per square meter;

Marginal value for **pc(stationary)** = $9.03395^{1.29649}$ (-0.01764) = -$316.39572 or -16.57% of sales price per square meter (minus sing here is an indication that coefficient on **pc(stationary)** goes with the minus sign) (Table 5.10).

Table 5.10: Marginal implicit prices, sales prices

		Linear model	Semi-log model	Box-Cox lhs
Estimated coefficient	**pc(mobile)** **pc(stationary)**	44.77610 -343.07580**	0.03380 -0.17520***	0.00898*** -0.01764***
Lambda		–	–	-0.29649***
Sample mean		1909.67800	1909.67800	1909.67800
Marginal value[1]	**pc(mobile)** **pc(stationary)**	44.77610 -343.07580**	64.54712 -334.57559***	161.06766*** -316.39572***
% of housing price	**pc(mobile)** **pc(stationary)**	2.34 -17.97**	3.38 -17.52***	8.43*** -16.57***

[1]MWTP = $\frac{\partial p}{\partial x_n} = p^{1-\lambda}\alpha_n x_n^{\lambda-1}$

Therefore, the obtained result suggests that a household in the apartment with mean sales price per square meter ($1909.68) would like to pay approximately $316.40 for each square meter of the apartment or 16.57% of the mean housing price to avoid a 13.5% increase in SO_2 (here we use the interpretation for the marginal change described above). Therefore, based on the analysis of sales housing prices, we conclude that residents of Kyiv on average would like to pay more for the reduction of air pollution that comes from stationary sources comparing to the pollution that comes from mobile emission sources.

Comparing the obtained results with the respective estimated from other countries, we do not find strong evidence that residents of Kyiv value clean air less than people from high-income countries. For example *Kim* et al. (2003) based on housing market for Seoul (Korea is considered as high-income economy by the World Bank) find that marginal willingness to pay for a 4% improvement in mean SO_2 concentrations constitute about 1.4% of mean housing price. In our case marginal willingness to pay for the decrease in SO_2 concentrations by 4% is around 4.9% of mean sales price (0.296 std change in **pc(stationary)** corresponds to 0.26 std change in SO_2). Such large difference can be attributed to the fact that marginal value of pollutants from stationary emission sources could be seen as the lower bound of the true effect due to the presence of some unobservable factors. As to the pollution from mobile emission sources, Kim et al. do not observe negative association with the housing prices, while in our case there is a significant positive relationship. Again, this may be due to the fact that marginal value of pollutants from mobile emission sources could be interpreted as the upper bound of the true marginal value. Therefore, we do not have strong evidence that people in Kyiv value clean air less than people in high-income countries; but in order to give complete answer to this question further analysis should be carried out.

For the sake of comparison of results obtained from analysis of rental and sales housing market data we present the respective estimates received from the model that includes only those variables that are present in both data sets. The same Box-Cox model with transformed dependent and independent variables is utilized. Tables 5.13 and 5.14 report the obtained results.

Table 5.13: Estimation results, Box-Cox model with transformed dependent and independent variables using the same parameter

	Rental prices	**Sales prices 1-room**	**Sales prices 2-room**	**Sales prices 3-room**	**Sales prices 4-room**
pc(mobile)	-0.02193***	0.00002	0.05554***	0.00976	-0.00235
	25.176	2.683	10.319	1.690	0.025
pc(stationary)	-0.07085***	-0.00016***	-0.11812***	-0.04931***	-0.01872
	109.077	32.596	25.164	22.086	1.221
area	-0.53346***	-0.00669***	0.15824***	0.1826***	0.10864***
	131.899	25.122	12.396	30.900	10.634
metro	-0.02967	-0.00001	0.01273*	0.01241***	-0.00107
	34.748***	0.697	3.626	13.043	0.017
park	0.05706***	-0.00003	-0.00427	0.02615**	0.01598
	81.803	1.388	0.056	5.050	1.189
Constant	2.73168	1.03673	5.44625	3.72897	2.57496
lambda	-0.30165***	-0.97014***	-0.06147	-0.16252**	-0.31610**
	(0.03830)	(0.09117)	(0.06359)	(0.06558)	(0.12477)

Table 5.13: Continued

mu	-0.30165*** 0.03830	-0.97014 (0.09117)	-0.06147 (0.06359)	-0.16252** (0.06558)	-0.31610** (0.12477)
Log likelihood	-3442.48810	-2152.96880	-3314.45400	-2735.52250	-952.13198
Observations	1688	302	451	367	122

LR statistic in parentheses
For transformation coefficients standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
All models include dummies for month and district

Table 5.14: Marginal implicit prices, rental and sales prices

		Rental prices	**Sales prices 1-room**	**Sales prices 2-room**	**Sales prices 3-room**	**Sales prices 4-room**
Estimated coefficient	**pc(m)**	-0.02193***	0.00002	0.05554***	0.00976	-0.00235
	pc(s)	-0.07085***	-0.00016***	-0.11812***	-0.04931***	-0.01872
Lambda		-0.30165***	-0.97014***	-0.06147	-0.16252**	-0.31610**
Sample mean		9.03395	1679.79900	1867.2400	1962.31700	2479.99400
Marginal value[1]	**pc(m)**	-0.21171***	45.21041	164.77324***	65.66891	68.94525
	pc(s)	-1.24325***	-361.68328***	-350.43240***	-331.77604***	549.21491
% of housing price	**pc(m)**	-2.34***	2.69	8.82***	3.35	-2.78
	pc(s)	-13.76***	-21.53***	-18.78***	-16.91***	-22.15

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

[1]MWTP $= \frac{\partial p}{\partial x_n} = p^{1-\lambda}\alpha_n x_n^{\lambda-1}$

Therefore, the marginal values for the pollution from stationary sources appear with the minus sign and are statistically significant in all cases, except for the 4-room apartments (this can be explained by the low number of observations); while marginal values for the pollution from mobile emission sources are negative and significantly different from zero for rental prices and go with positive sign for 1-, 2-, and 3-room apartments. So, we again come to the conclusion that people in Kyiv on average would like to pay more for the reduction of air pollution that comes from stationary sources comparing to the pollution that comes from mobile emission sources since the effect of pollution from mobile emission sources on housing prices is not so clear (it is negative for the rental prices and positive and insignificant for the 1-room apartments).

This finding also supports the idea that rental and sales housing markets in Kyiv can differently respond to some factors.

Here we also test for the existence of market segmentation. According to the *Freeman* (2003) the market could be considered as segmented if buyers have different preferences or housing characteristics are different across submarket and there are some informational or geographical barriers between submarkets. If the market under investigation is really segmented, separate hedonic price equations should be estimated for each submarket since non taking into account the segmentation issue could lead to faulty estimates of the marginal implicit prices (Freeman, 2003). We can hypothesize that housing market of Kyiv is segmented due to the geographical barrier, i.e. division of the urban area by the Dnipro River. In order to test this hypothesis, ideally, we should estimate separate hedonic price equation for the right- and left-bank parts of the city and then compare the obtained results. Unfortunately, we can not follow this procedure due to the specific features of out data (we are dealing with the sub-samples of houses and not with the whole sample). But we can still make some inferences by incorporating interactions of dummy indicating left-bank part of the city with the principal components reflecting air pollution (**left_pc(mobile)** and **left_pc(stationary)** variables). If the coefficients on these interactions are statistically significant from zero, this may be an indication of the existence of market segmentation. Table 5.12 reports the estimation results both for rental and sales housing markets.

Table 5.12: Estimation results, Box-Cox models, testing for the market segmentation

	Box-Cox model lhs+rhs[1] rental prices	Box-Cox model lhs sales prices
pc(mobile)	0.00224 (0.18000)	0.00388** (4.00500)
pc(stationary)	-0.12112*** (206.16400)	-0.01812*** (35.78000)
left_pc(mobile)	-0.02276 (0.46200)	-0.01702* (2.70100)
left_pc(stationary)	0.3985*** (93.03200)	0.02762 (0.87400)
area	-0.70633*** (406.04800)	(0.00013)*** (12.63000)
room1	–	0.01504** (9.89400)
room2	–	0.01056** (6.14200)
room3	–	0.00289 (0.58900)
floor1	–	-0.01115*** (14.76200)
floor2	–	-0.01345*** (25.53100)
wall	–	0.00095 (0.09900)
height	0.03501 (2.63200)	–
balcony	0.00018 (0.00100)	–
door	0.00659 (0.92200)	–
tel	0.0005 (0.00400)	-0.00638** (10.33700)
tv	0.02395*** (12.78800)	–
appliances	0.08971*** (91.88500)	–
furniture	0.01057 (1.57600)	–
renovated	0.04678*** (45.33700)	–

Table 5.12: Continued

Euroremont	0.22677*** (257.95700)	–
studio	0.05871** (5.15500)	–
metro	-0.04432*** (46.91100)	0.00004** (11.12400)
park	-0.01466 (1.74000)	0.00720** (6.05000)
Constant	3.67189	2.975301
lambda	-0.17788*** (0.03473)	-0.30008*** (0.02680)
mu	-0.17788*** (0.03473)	–
Log likelihood	-3161.54170	-8597.78440
Observations	1688	1133

LR statistic in parentheses
For transformation coefficients standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
The models include dummies for month and district
[1]lhs=transformation of the left-hand side (dependent variable)
rhs=transformation of the right-hand side (independent variables)

Coefficient on **left_pc(mobile)** is negative both for rental and sales housing market data but it is not statistically significant from zero at 5% level. Coefficient on **left_pc(stationary)**, on the other side, goes with the positive sign and is statistically significant at 5% level for the rental housing market data. Therefore, we do not find strong evidence of housing market segmentation in Kyiv due to the geographical barrier, which is an indication that our previous analysis is valid. But further analysis of this issue could be carried out.

CONCLUSION

This paper provides estimation of marginal willingness to pay for the air quality improvement using hedonic price analysis. It is conducted for housing market in Kyiv and is motivated by the high air pollution in Kyiv on the one hand, and lack of applications of hedonic price techniques for valuation environmental amenities in transition countries on the other hand. This study utilized data on rental and sales housing markets with combination of the data on four air pollutants: particulate matter (PM), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), and carbon monoxide (CO).

It appears that residents of Kyiv are concerned more about pollution that comes from stationary sources comparing to pollution that come primary from mobile emission sources, which is indicated by their positive willingness to pay for improvement in SO_2 concentrations. This relationship is supported in the analysis of both rental and sales housing market data: marginal willingness to pay for a 13.5% decrease in mean SO_2 concentrations is about \$1.01 for each square meter of the apartment or 11.20% of mean rental price and \$316.40 or 16.57% of mean sales price. On the other hand, results for marginal willingness to pay for improvement in ambient air concentrations of pollutants from mobile emission sources are somewhat different for the rental and sales housing market data, which also reinforce the idea that rental and sales housing markets in Kyiv are different.

However, we should very careful interpret the estimated marginal effects. Marginal value of pollutants from mobile emission sources could be interpreted as the upper bound of the true marginal value, while marginal value of pollutants from stationary emission sources could be seen as the lower bound of the true effect due to the presence of some unobservable factors. The possible influence of omitted factors does not allow comparing the obtained results with the respective estimates from other countries, but even rough comparison shows that there is no strong evidence that people in Kyiv value clean air less than people in high-income countries.

This research has important policy implications:

- The obtained econometric results suggest that residents in Kyiv do value air quality as indicated by the positive willingness to pay for improvement in SO_2 concentrations. Therefore, any attempts by the local government to improve the air quality will be associated with the increase in welfare to the people;
- Moreover, people in Kyiv are concerned more about pollution that comes from stationary sources comparing to pollution that come primary from mobile emission sources. This is also consistent with the Ukrainian legislation. For example, in the Decision of Cabinet of Ministers of Ukraine about the use of budget funds for the increase in air quality in 2007, primary attention is given to the pollution from stationary sources. Given the fact that approximately 70% of air pollution in Kyiv comes from transportation (Central Geophysical Observatory) and its rapid growth rates, the local government should pay more attention to this issue both by disseminating information about this problem and by improving legislation.

Moreover, real estate specialists could use the obtained results while estimating the value of apartments by explicitly incorporating the environmental component in the value of apartments.

In conclusion, this research is the first attempt to assess the economic value that residents of Kyiv place on clean air suggesting that they indeed care about air quality. Although current study suffers from several limitations, it provides a good basis for further analysis.

BIBLIOGRAPHY

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Kluwer, Dordrecht.

Boyle, M. and K. Kiel. 2001. *A Survey of House Price Hedonic Studies of the Impact of Environmental Externalities*. Journal of Real Estate Literature. Vol. 9, No 2, pp. 117-144.

Chartfiled, C. and A. Collins. 1980. *Introduction to Multivariate Data Analysis*. London: Chapman and Hall, pp. 57-79.

Chattopadhyay, S. 1999. *Estimating the Demand for Air Quality: New Evidence Based on the Chicago Housing Market*. Land Economics, Vol. 75, pp. 22-38.

Cropper, M., L. Deck and K. McConnell. 1988. *On the Choice of Functional Form for Hedonic Price Functions*. The Review of Economics and Statistics, Vol. 70, No. 4, pp. 668-675.

Ekeland, I., J. Heckman and L. Nesheim. 2002. *Identifying Hedonic Models*. The American Economic Review, Vol. 91, No. 2, pp. 304-309.

Freeman III, A. 2003. *Property Value Models*. Chapter 11 in *The Measurement of Environmental and Resource Values*: *Theory and Methods*. Washington: RFF Press, pp. 353-398.

Green, W. 4th edition. *Econometric Analysis*. Prentice-Hall, pp.444-455.

Haab, T., and K. McConnell. 2002. *Hedonic Price Equations*. In *Valuing Environmental and Natural Resources: the Econometrics of Non-market Valuation.* Cheltenham, Northampton: Edward Elagar, pp. 245-267.

Halvorsen, R. and H. Pollakowski. 1981. *Choice of Functional Form for Hedonic Price Equations*. Journal of Urban Economics, Vol. 10, No. 1, pp. 37-49.

Harrison, D. and D. Rubinfeld. 1978. *Hedonic Housing Prices and the Demand for Clean Air*. Journal of Environmental Economics and Management, Vol. 5, pp. 81-102.

Kim, C., T. Phipps and L. Anselinc. 2003. *Measuring the Benefits of Air Quality Improvement: a Spatial Hedonic Approach*. Journal of Environmental Economics and Management, Vol. 45, pp. 24-39.

Lancaster, K. 1966. *A New Approach to Consumer Theory*. The Journal of Political Economy, Vol. 74, No. 2, pp. 132-157.

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

Murty, M., S. Gulati and A. Banerjee. 2003. *Hedonic Property Prices and Valuation of Benefits from Reducin Urban Air Pollution in India.* Institute of Economic Growth, Delhi University Enclave, pp. 1-27.

Rosen, S. 1974. *Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition*. Journal of Political Economy, Vol. 82, pp. 34-55.

Sioma, D. 2006. *Rental Housing Price Differentials and the Cost of Commuting. Evidence from Kyiv.* EERC MA Thesis.

Smith, V. and J. Huang. 1995. *Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models*. Journal of Political Economy, Vol. 103, pp. 209-227.

Strukova, E., A. Golub and A. Markandya. 2006. *Air Pollution Costs in Ukraine.*

Yusuf, A. and B. Resosudarmo. 2006. *Assessing the Value of Clean Air in a Developing Country: A Hedonic Price Analysis of the Jakarta Housing Market, Indonesia.* Australian National University, York, USA.

Economics and Environment Network Working Paper, pp. 1-24.

World Bank. 1992. *World Development Report 1992.* Oxford University Press, New York, USA.

APPENDIX

APPENDIX A. DESCRIPTION OF AIR POLLUTION

Table A1: Description of Air Pollution

Pollutant	Station, №	Number of observations	Annual averarage concentr., mg/m^3	Maximum		Frequency, %		Index of pollution
				Observable, mg/m^3	Month	>MAC	>5MAC	
PM	01	594	0.0856	0.1998	10	0.0	0.0	0.57
	02	570	0.0935	0.1998	11	0.0	0.0	0.62
	03	606	0.0970	0.1998	11	0.0	0.0	0.65
	04	558	0.1406	0.1998	12	0.0	0.0	0.94
	05	532	0.0806	0.1998	9	0.0	0.0	0.54
	06	527	0.1026	0.1998	12	0.0	0.0	0.68
	07	593	0.1514	0.1998	12	0.0	0.0	1.01
	08	606	0.1391	0.1998	12	0.0	0.0	0.93
	09	516	0.1061	0.1998	12	0.0	0.0	0.71
	10	590	0.1325	0.1998	12	0.0	0.0	0.88
	11	608	0.1672	0.1998	12	0.0	0.0	1.11
	13	606	0.0730	0.0998	12	0.0	0.0	0.49
	15	568	0.0881	0.1998	12	0.0	0.0	0.59
	17	591	0.0991	0.1998	11	0.0	0.0	0.66
	20	602	0.1194	0.1998	12	0.0	0.0	0.80
	21	232	0.1206	0.1998	5	0.0	0.0	0.80
SO_2	01	1188	0.0126	0.0458	2	0.0	0.0	0.25
	02	1140	0.0141	0.0458	12	0.0	0.0	0.28
	03	606	0.0106	0.0398	1	0.0	0.0	0.21
	04	606	0.0118	0.0498	9	0.0	0.0	0.24
	05	1072	0.0103	0.0438	2	0.0	0.0	0.21
	06	509	0.0125	0.0628	5	0.0	0.0	0.25
	07	606	0.0123	0.0428	2	0.0	0.0	0.25
	08	544	0.0132	0.0418	1	0.0	0.0	0.26
	09	499	0.0132	0.1258	6	0.0	0.0	0.26

	10	590	0.0130	0.1168	6	0.0	0.0	0.26
	11	1216	0.0163	0.1298	6	0.0	0.0	0.33
	13	1212	0.0114	0.0398	5	0.0	0.0	0.23
	15	1136	0.0127	0.0688	2	0.0	0.0	0.25
	17	1212	0.0165	0.0548	12	0.0	0.0	0.33
	20	1204	0.0165	0.0518	12	0.0	0.0	0.33
	21	219	0.0154	0.0408	1	0.0	0.0	0.31
CO	01	594	1.6515	9.0000	7	1.5	0.0	0.60
	02	570	1.7350	9.0000	6	1.4	0.0	0.63
	03	606	1.2673	7.0000	10	0.7	0.0	0.48
	04	606	1.4405	7.0000	6	0.8	0.0	0.54
	05	532	0.7311	2.0000	3	0.0	0.0	0.30
	06	527	2.4402	10.0000	5	4.4	0.0	0.84
	07	606	2.8168	24.0000	1	7.4	0.0	0.95
	08	606	1.8399	8.0000	10	1.5	0.0	0.66
	09	516	2.5852	15.0000	4	8.7	0.0	0.88
	10	606	2.0148	10.0000	7	3.8	0.0	0.71
	11	608	2.1249	8.0000	1	3.0	0.0	0.75
	13	606	0.8300	3.0000	9	0.0	0.0	0.34
	15	568	1.0580	7.0000	8	0.2	0.0	0.41
	17	606	1.3283	8.0000	8	0.3	0.0	0.50
	20	606	2.6419	9.0000	5	5.3	0.0	0.90
	21	464	0.8749	4.0000	6	0.0	0.0	0.35
NO_2	01	1188	0.1177	0.2898	5	73.1	0.0	4.07
	02	1140	0.1194	0.2698	5	73.6	0.0	4.14
	03	606	0.1076	0.3598	5	64.0	0.0	3.62
	04	606	0.1042	0.2500	7	60.4	0.0	3.47
	05	1072	0.0519	0.2298	11	15.8	0.0	1.40
	06	509	0.1023	0.4598	5	53.6	0.2	3.39
	07	606	0.1326	0.9798	5	74.8	0.8	4.75
	08	544	0.1219	0.7398	7	67.3	0.2	4.25
	09	499	0.1145	0.3498	5	65.5	0.0	3.92
	10	590	0.1109	0.4398	5	64.1	0.2	3.76
	11	1216	0.1202	0.2998	5	73.2	0.0	4.18
	13	1212	0.0501	0.2198	5	13.9	0.0	1.34
	15	1136	0.0703	0.2298	11	27.6	0.0	2.08
	17	1212	0.1244	0.7098	7	74.0	0.2	4.37
	20	1204	0.1093	0.3998	5	70.1	0.0	3.69
	21	219	0.0850	0.1898	5	42.9	0.0	2.66

APPENDIX B. ESTIMATION PROCEDURE FOR THE BOX-COX MODEL

The estimation procedure for the model $p_i(\lambda) = \alpha_0 + \sum_{j=1}^{n} \alpha_j x_{ij}(\lambda) + \varepsilon_i$ is shown below based on explanations provided in *Green* (2000). Assuming that $\varepsilon \sim N(0,\sigma^2)$, the density for ε_i is given by $f(\varepsilon_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right)$, and the log-likelihood function for N observations is the following

$$\ln L = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\varepsilon_i^2 .$$

In our case $\varepsilon_i = p_i(\lambda) - \alpha_0 - \sum_{j=1}^{n} \alpha_l x_{ij}(\lambda)$, and the density is

$$f\left(p_i(\lambda) - \alpha_0 - \sum_{j=1}^{n} \alpha_l x_{ij}(\lambda)\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(p_i(\lambda) - \alpha_0 - \sum_{j=1}^{n} \alpha_l x_{ij}(\lambda)\right)^2}{2\sigma^2}\right).$$

To find the probability density function for p_i we can use the following transformation: $f(p_i) = \left|\frac{d\varepsilon_i}{dp_i}\right| f(\varepsilon_i)$. So, the probability density function for p_i can be written as $f(p_i) = p_i^{\lambda-1} f\left(p_i(\lambda) - \alpha_0 - \sum_{j=1}^{n} \alpha_l x_{ij}(\lambda)\right)$.

The log-likelihood function for the linear Box-Cox model is

$$\ln L = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\sigma^2 + (\lambda - 1)\sum_{i=1}^{N}\ln p_i - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(p_i(\lambda) - \alpha_0 - \sum_{j=1}^{n} \alpha_l x_{ij}(\lambda)\right)^2 .$$

The concentrated log-likelihood function is given by

$$\ln L_c = (\lambda - 1)\sum_{i=1}^{N} \ln p_i - \frac{N}{2}\left(\ln(2\pi) + 1\right) - \frac{N}{2}\ln \hat{\sigma}^2,$$

where we used the result that the estimator of σ^2 was the average squared residuals. By maximizing the concentrated log-likelihood function we can find the estimates of λ and α's. Since the least squares standard errors always underestimate the correct asymptotic standard errors, we need to correct them. We will follow the procedure presented in *Green* (2000) to compute the estimator for the asymptotic covariance matrix for the maximum likelihood. For the Box-Cox model

$$\ln f(p_i) = (\lambda - 1)\ln p_i - \frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\left(p_i(\lambda) - \alpha_0 - \sum_{j=1}^{n} \alpha_l x_{ij}(\lambda)\right)^2.$$

Taking the derivatives

$$\frac{\partial \ln f(p_i)}{\partial \alpha} = \frac{\varepsilon_i}{\sigma^2}\mathbf{x}_i(\lambda),$$

$$\frac{\partial \ln f(p_i)}{\partial \lambda} = \ln p_i - \frac{\varepsilon_i}{\sigma^2}\left(\frac{\partial p_i(\lambda)}{\partial \lambda} - \sum_{l=1}^{n} \alpha_j \frac{\partial x_{ij}(\lambda)}{\partial \lambda}\right),$$

$$\frac{\partial \ln f(p_i)}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{\varepsilon_i^2}{2\sigma^4}$$

we can express the estimator of the asymptotic covariance matrix of the parameters estimates as Est. Asym. $\mathrm{Var}(\hat{\theta}) = \left(\sum_{i=1}^{N} \hat{\mathbf{w}}_i \mathbf{w}'\right)^{-1}$ were $\mathbf{w}_\mathrm{i}$ if given by

$$\mathbf{w}_i = \begin{pmatrix} \partial \ln f(p_i)/\partial \alpha \\ \partial \ln f(p_i)/\partial \lambda \\ \partial \ln f(p_i)/\partial \sigma^2 \end{pmatrix}$$

APPENDIX C. ESTIMATION RESULTS

Table C1: Estimation results, Box-Cox model with transformed dependent and independent variables using the same parameter, rental prices per square meter as dependent variable

pc(mobile)	-0.02224*** (22.51400)	**pc(stationary)**	-0.07172*** (100.85600)
area	-0.69215*** (347.76700)	**metro**	-0.03306*** 35.611
height	0.03796* (3.18800)	**park**	0.06211*** (77.55600)
balcony	0.00018 (0.00100)	**district1**	-0.02062 (1.21100)
door	0.00368 (0.29700)	**district2**	-0.1979*** (55.45800)
tel	0.00188 (0.05600)	**district3**	-0.16245*** (43.57100)
tv	0.0218*** (10.93200)	**district4**	-0.05964*** (10.39800)
appliances	0.09277*** (101.78800)	**district6**	0.12295*** (48.02900)
furniture	0.00229 (0.07700)	**district7**	0.01153 (0.49700)
renovated	0.04862*** (50.54900)	**district8**	0.0944*** (25.62900)
euroremont	0.22226*** (256.26200)	**district9**	-0.02933 (1.93300)
studio	0.05594** (4.82100)	**district10**	-0.04459** (5.60500)
Constant	3.41099		
lambda	-0.20248*** (0.03505)		
mu	-0.20248*** (0.03505)		
Log likelihood	-3170.22890		

Observations 1688
LR statistic in parentheses
For transformation coefficients standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
The model includes dummies for month

APPENDIX D. ANALYSIS OF HOUSING SALES PRICES

Using test proposed by MacKinnon, White, and Davidson we test linear model versus semi-log model (Table D.1). Coefficient on new variable $\ln \hat{p}_i f - \ln \hat{p}_i$, where $\hat{p}_i f$ are the fitted values from linear model and $\ln \hat{p}_i$ are the fitted values from semi-log model, is -1509.0940 with t-statistic -4.1200. Thus, we reject the null hypothesis of linear functional form in favour of semi-log model.

Table D1: Estimation results, linear and semi-log regression models, sales prices pre square meter as dependent variable

	Linear model	**Semi-log model**	**Semi-log model with some variables squared**	**Semi-log model without influential observations**
pc(mobile)	44.7761 (-89.9043)	0.0322* (-0.0193)	0.0338 (-0.0210)	0.0352** (-0.01620)
pc(stationary)	-343.0758** (-139.6440)	-0.1658*** (-0.0300)	-0.1752*** (-0.0306)	-0.1588*** (-0.02520)
room1	442.7402** (-222.3920)	0.1547*** (-0.0477)	0.1544*** (-0.0528)	0.1403*** (-0.04010)
room2	209.9826 (-198.1090)	0.1038** (-0.0425)	0.0995** (-0.0442)	0.0978*** (-0.03570)
room3	64.7244 (-176.0460)	0.0330 (-0.0378)	0.0289 (-0.0378)	0.0417 (-0.03170)
area	2.8593 (-1.7606)	0.0014*** (-0.0004)	0.0015 (-0.0011)	0.0016*** (-0.00030)
areasq	–	–	-5.40e-07 (3.51e-06)	–
floor1	-73.9896 (-134.8750)	-0.0922*** (-0.0289)	-0.0912*** (-0.0290)	-0.0796*** (-0.02450)
floor2	-363.6756*** (-123.4750)	-0.1372*** (-0.0265)	-0.1403*** (-0.0266)	-0.1329*** (-0.02220)
tel	-210.8628** (-92.1952)	-0.0701*** (-0.0198)	-0.0665*** (-0.0200)	-0.0667*** (-0.01670)
wall	56.6401 (-140.6890)	0.0130 (-0.0302)	0.0078 (-0.0303)	0.0112 (-0.02540)

Table D1: Continued

metro	2.0447*** (-0.6192)	0.0005*** (-0.0001)	-0.0209 (-0.0138)	0.0005*** (-0.00010)
metrosq	–	–	0.0001 (3.36e-05)	–
park	80.1174 (-108.8450)	0.0415* (-0.0234)	-0.0661 (-0.1158)	0.0457** (-0.01970)
parksq	–	–	0.0341 (-0.0350)	–
Constant	1,943.5781*** (-367.9740)	7.4892*** (-0.0790)	7.5658*** (-0.1174)	7.4917*** (-0.06640)
Observations	1133	1133	1133	1125
R-squared	0.29	0.63	0.63	0.71

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

We check whether there is a linear relationship between sales prices (to be precise, sales prices per square meter) and area, and also between rental prices and variables that reflect proximity to the metro station and green spaces. Including area squared (**areasq**) into the semi-log model, we find that coefficient on this variable is not statistically significant at 5% level. Coefficient on **areasq** equals to 5.40e-07 with t-statistic -0.1500. Coefficients on **metrosq** (0.0001with t-statistic 1.5500) and **parksq** (0.0341with t-statistic 0.9700) are not statistically significant at 5% level either. Therefore, we will not include area squared and variables that reflect proximity to the metro station and green spaces squared into our further regression analysis.

We check the semi-log model for misspecification and robustness by finding influential observations first. To identify influential cases we use Cook's Distance. Figure D.1 shows that observations 142, 166, 208, 428, 542, 665, 1100, and 1250 have large Cook's distances. To find the reasons for influence we look at outliers (Figure D.2) and leverage (Figure D.3).

Figure D.1: Cook's Distance from the semi-log model, measure of overall influence

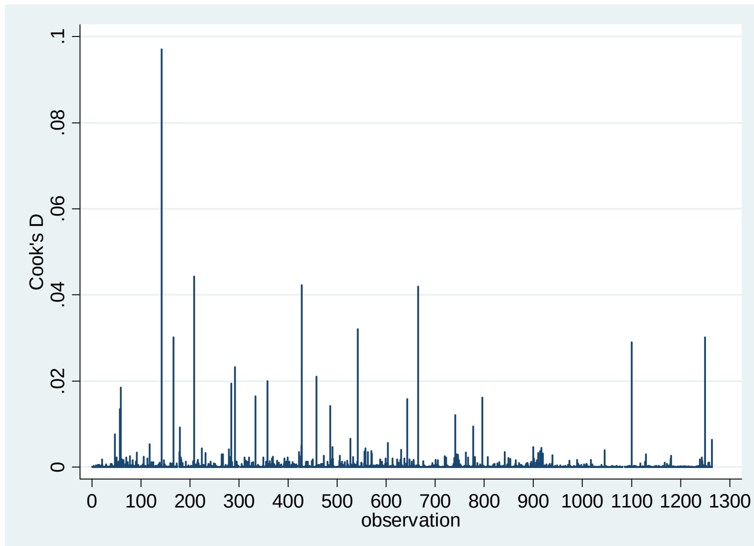
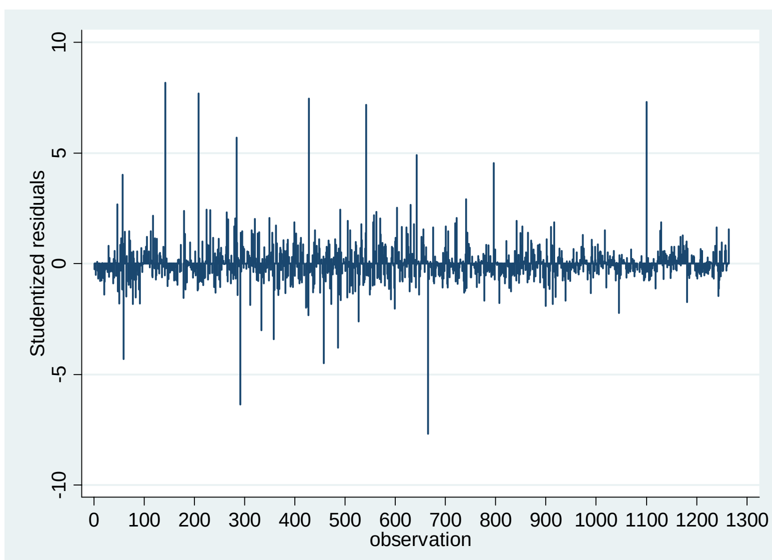
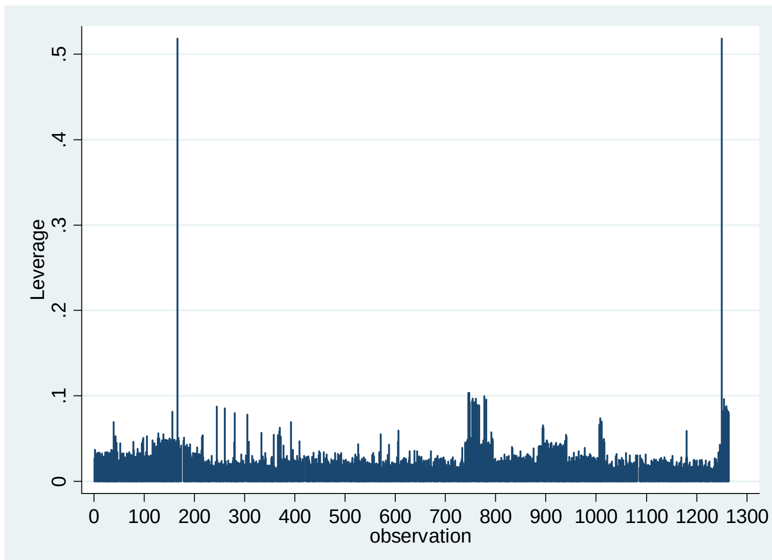


Figure D.2: Studentized residuals for detecting outliers from the semi-log model



Large studentized residuals for observations 142, 208, 428, 542, 665, and 1100 (Figure D.2) indicate that these observations are outliers. Observations 166 and 1250 have very large leverage (Figure D.3). Therefore, we exclude observations 142, 166, 208, 428, 542, 665, 1100, and 1250 from our sample.

Figure D.3: Leverage from the semi-log model



Checking the semi-log model for misspecification and robustness, we also find that the model suffers from non-normality and heteroskedasticity. The probability that residuals from semi-log model normally distributed is less than 0.0010 based on Skewness/Kurtosis test for normality. We also reject the null hypothesis that residuals are homoskedastic (i.e. they have constant variance) on the basis of the Breusch-Pagan/Cook-Weisberg test. $\chi^2(1)$-statistics equals to 81.97.

We use the procedure described above to test whether we need to transform dependent variable or continuous independent variables. We estimate that coefficient on a_i equals 0.9868 and is statistically significant from zero at 5% level (Table D.2). Thus, we need the transformation of dependent variable. To check whether we need to transform the continuous independent variables (**area, metro, park**) we include into the linear model additional variables:

area_at = area $*$ log(area)

metro_at = metro $*$ log(metro)

park_at = park $*$ log(park)

and test whether the estimated coefficients on this variables are significant (Table D.2). The coefficient on **area_at** equals to -0.6382, on **metro_at** it is 7.4142, on **park_at** it is -17.3348. All coefficients are not statistically significant at 5% level. This is an indication that we should not transform independent variables. Therefore, we conclude that we need to transform only dependent variable (sales prices per square meter).

Table D.2: Estimation results, linear model, sales prices per square meter as dependent variable, testing for the need of variables transformation

	Linear model **Test for the need of dependent variable transformation**	**Linear model** **Test for the need of independent variables transformation**
pc(mobile)	53.6145** (-23.0097)	44.6150* (-25.2581)
pc(stationary)	-240.8328*** (-35.7587)	-245.8385*** (-36.8235)
room1	179.2441*** (-56.9491)	193.3785*** (-65.1728)
room2	159.3917*** (-50.6957)	159.5788*** (-52.6494)
room3	77.4943* (-45.0307)	72.0459 (-45.0747)
area	1.7238*** (-0.4538)	5.3699 (-6.1339)
floor1	-132.6870*** (-34.7914)	-132.0942*** (-34.8025)

Table D.2: Continued

floor2	-173.5227*** (-31.7115)	-174.9355*** (-31.7811)
tel	-78.3596*** (-23.8209)	-72.9186*** (-24.0762)
wall	20.1343 (-36.0158)	16.2413 (-36.1518)
metro	0.6057*** (-0.1594)	-44.0507* (-22.9908)
park	44.3627 (-27.9828)	72.7537 (-138.704)
a_i	0.9868*** (-0.0180)	0.9888*** (-0.0181)
area_at	–	-0.6382 (-1.0482)
park_at	–	-17.3348 (-99.2837)
metro_at	–	7.4142* (-3.8174)
Constant	3,469.4615*** (-98.7102)	3,415.4363*** (-196.472)
Observations	1125	1125
R-squared	0.89	0.89

Standard errors in parentheses

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