



LUND UNIVERSITY
School of Economics and Management

Duration on the Housing Market: A Nonparametric Approach

Yana Petrova¹

Supervisor: Peter Jochumzen

Lund University

Department of Economics

Master Essay – II (NEKP01)

2nd June 2015

¹ yana.d.petrova@gmail.com

Abstract

This study examines the relationship between the time-on-the-market and the price of apartments on the housing market in Moscow. It is based on a data set including over 10 000 property objects during the time period January 2012 - October 2014. Only secondary property is considered. The problem of right-censoring is present in the sample, which means that it is not possible to use the duration data without preliminary adjustment. A nonparametric approach is employed to estimate the duration for the observations which are censored. First, an estimation of the hazard rate function is obtained. Second, based on the hazard rate the values of censored durations are restored with the help of simulation techniques. The duration data adjusted for censoring is used to examine if the market duration depends on the housing price and type of the seller. The duration data shows to exhibit a positive duration dependence. For the Moscow property market, a positive relationship between price and time-on-the-market is found. On average it takes longer time to sell the property with the help of real estate agents than without it.

Keywords: Housing Market, Duration, Censoring, Nonparametric approach, Statistical simulations, Bootstrap, Kaplan-Meier estimator

Table of Contents

1. INTRODUCTION.....	4
2. LITERATURE REVIEW.....	6
2.1. BUYERS.....	7
2.2. SELLERS.....	9
3. DATA.....	12
4. METHODOLOGY.....	14
4.1. ESTIMATED HAZARD FUNCTION.....	14
4.2. SIMULATION PROCEDURES TO OBTAIN THE PARAMETERS OF INTEREST.....	16
5. EMPIRICAL ANALYSIS.....	19
5.1. PRELIMINARY ANALYSIS.....	19
5.2. DURATION SIMULATIONS.....	21
5.3. ESTIMATION RESULTS.....	23
6. CONCLUSIONS.....	25
7. REFERENCES.....	26
8. APPENDIX.....	28
8.1. A1. TABLE: DESCRIPTION OF THE VARIABLES.....	28
8.2. A2. GRAPH: KAPLAN-MEIER ESTIMATOR OF THE SURVIVAL FUNCTION.....	29
8.3. A3. GRAPH: LOGARITHMS OF MARKET DURATION AND PRICE.....	29

1. Introduction

Ever since property rights were established, people have been selling houses, thus, the real estate market was formed far back in the past. Even though the market itself has existed for a long time, economists started to consider it as an object of research only about half a century ago. In the beginning, the classical demand-supply approach was applied when analysing the housing market. However, there is at least one very specific characteristic of the market that makes it less typical and harder to analyse. Specifically, a house is a durable good, which means that the buyer carefully approaches the question of buying it and takes his/her time to find a perfect deal. This means that a property unit as an object of sale can stay on the market for a while before actually being sold. So such a characteristic as time-on-the-market (TOM) is important when describing the real estate market.

Time-on-the-market is defined as the number of days, weeks or any other time units between the days when the object entered the market until it is sold. This term can be also referred to as the duration (usually used in the empirical analysis) or marketing time.² Viewing the housing market from the perspective of duration, instead of applying the classical demand-supply point of view, changes the approach to a market analysis. When looking at the duration on the market it is no longer of interest that the price is formed by the demand and supply sides. Instead, it is supposed that prices are given by sellers and buyers choose among the houses available on the market, which implies that the duration is a measure of search costs. When it comes to analysing duration different aspects are of interest. Duration dependence is one of them, which could potentially be both positive and negative. For instance, a positive duration dependence implies that the probability of sale increases the longer the object is on the market, while a negative duration dependence is a sign of the opposite, which could also be possible in some circumstances. This question is often studied by estimating the survival function, a method which will be discussed later. However, the main focus of the current study is on the relationship between market duration and price.

The question of which factors influence the time-on-the-market has been of specific interest to the researchers in the recent years. Different ones have been considered, among which price plays a crucial role. In particular, the matter here is if a high price increases the TOM as people are less likely to buy expensive objects, or on the contrary, if a high price serves as a signal of the property being “exclusive” and, thus, helps to sell the object quicker. In this

² The term “marketing time” is considered, because often the listing date and the date of sale are considered to be the ones when owner started to advertise the property till the advertisement was closed.

case it is important to clarify the side factors which influence the housing prices and make them incomparable before trying to analyse the relationship between price and market duration. For example, such factor as location: it could be that two identical apartments are situated in different areas and thus have different prices, however this does not mean that the more expensive object is likely to stay longer on the market. It could actually happen that the opposite is true, so in this case location difference makes prices incomparable and, thus, needs to be corrected for. Furthermore, there is another problem occurring when it comes to analysing duration data, the issue of censored observations being present in the data.

Although analysing TOM attracts a lot of attention from researchers, the duration data often comprises a right-censoring issue, which causes problems when it comes to the estimation part. In particular, censored data imply that it is not possible to obtain reliable results before dealing with the problem. Modern research in the housing market field relies on parametric methods when it comes to dealing with censored data. However, the main disadvantage of the parametric approach is that it requires strong distributional assumptions. It is of crucial importance that the distribution assumption is accurate, otherwise the obtained estimates have no power. Though the advantage is that this approach allows for the censoring problems in the data without actually dealing with them. An alternative approach to coping with the issue mentioned above is a nonparametric one. Particularly, it allows modelling the hazard function based on the data without any distributional assumptions. Even though the nonparametric method has an apparent advantage, it has not been applied to the duration analysis on the housing market, to the knowledge of the author. Thus, the current study contributes to the previous research by suggesting an application of the named above method.

The focus of the current study is on applying a nonparametric approach when dealing with censored duration data in order to examine the relationship between the TOM and price on the housing market. The nonparametric approach is considered in order to obtain numerical estimates of the hazard function. Based on those values the duration for censored observations is modelled. A simulation technique is used to obtain the estimations of the duration values for the observations which do not have a complete TOM in the available sample. Then the relationship between TOM and housing price is investigated, other possible factors influencing the market duration are considered as well.

Most of the duration research is carried out for the United States real estate market and to a smaller extent economists also consider different European markets. However, the Russian property market has not earlier been analysed from a time-on-the-market perspective. In light of the above, the purpose of the current study is not only to consider a method that has not been

previously applied to the housing market, but also to examine a housing market which is yet unexplored from the duration perspective.

A data set for the Moscow housing market has been obtained from an online listing portal “www.gdeetotdom.ru”. The sample contains apartments that are resold (newly built housing is not taken into account). The advantage of analysing the secondary market is that it contains a larger variation of selling conditions as then property objects are relatively more diverse compared to newly built properties, which allows to check which particular characteristics influence the TOM to a higher extent. The size of the sample is an advantage of the current study. Over 10 000 property objects are considered. This makes it possible to argue for the quality of the obtained results, meaning that if a relationship between the variables of interest is found, then the conclusion can be extended to the Moscow housing market as a whole, allowing the comparison with other housing markets around the globe.

The remainder of this paper is organized as follows: the second section gives an overview of the most relevant literature in the field of duration analysis on the housing market. The third section describes the data set, paying attention to the source of the data. The fourth section describes the methodology for duration analysis and simulation technique. The following section provides the estimation results, while the final section concludes.

2. Literature Review

The existence of housing markets goes far back in time, however, only in the past half a century it has become an object of interest to researchers. Smith et al. (1988) point out that the economic literature in the sphere of real estate evolved extensively during 1970-1980s. They point out that when housing had been recognised not only as a commodity which responds to market forces, but also as an object with a number of specific characteristics which require that the estimation models are modified in order to conduct adequate analysis.

Different aspects of housing markets are considered in the economic literature. A lot of economists focus on examining the factors forming demand and supply on the real estate market and the determinants of the property value, as well as the quantities being sold. For example, Chay and Greenstone (2005) exploit the impact that air pollution has on property values. Based on a microdata set for the period 1970-1990 they conclude that, *ceteris paribus*, housing prices tend to rise as pollution levels fall. De Wit et al. (2012) look at the correlation between the rate of price increase and the number of houses sold. They consider the Dutch housing market during the period 1985-2007 and estimate a vector error correction model, which provides strong

argumentation for the existence of high correlation between the considered variables. Paying attention only to the demand side Jaén-García and Piedra-Muñoz (2012) examine the Spanish housing market in 2006-2007. A probit model is used to analyse the probability of housing purchase and a model for consumption is also estimated, based on which reactions of demand to fiscal condition changes are examined. Díaz and Jerez (2013) focus on the cyclical properties of the key macroeconomic time series of the housing market, such as volatility of prices and sales with respect to GDP and some others. They calibrate a model with aggregate demand and supply on the United States real estate market in order to carry out an analysis of cyclical properties. They also pay attention to the fact that the average time-on-the-market is highly volatile and negatively correlated with houses prices.

Even though there is still a lot of research being done in the classical areas of market analysis, the TOM is another characteristic of housing markets that is often analysed as being important for exactly this type of market. Time-on-the-market has been considered as a subject of analysis both on its own and in relation to housing prices and other variables. Different factors influencing the duration time have been examined, varying from most expected ones such as price of the house or its size and location to more atypical ones. For instance, Benefield et al. (2011) test if the property photo depiction in a multiple listing service (MLS) affects property price and time-on-the-market.

If now focusing on more typical aspects of TOM for housing, it should be noted that there are two possible points of view on the topic, as there are two agents involved in the house-selling process: a buyer and a seller. Even though the current research focuses on the sellers side, the literature referring to both sides is considered in order to provide the reader with an overall picture of the research area. To begin with, the studies focusing on buyers' side are briefly discussed. Then a discussion on sellers' side will be provided in more details.

2.1. Buyers

To the knowledge of the author, the first articles related to TOM for real estate belong to the early 1960s. One of the first to pay attention to this question was Stigler (1961). He looks at the buyers search process and examines the relationship between price and duration. The search process is defined as follows: the buyer who wishes to obtain the most favourable price must first identify the property objects he/she is interested in and then seek out prices from the sellers (Stigler 1961). Such a process is time-consuming. Stigler concludes that increased time

of search results in finding a lower priced house, meaning that the asking price³ is negatively related to duration. In this case the duration is measured as the time span, during which the buyer is searching for a house to purchase. The author, however, assumes that the buyers are involved in the search process directly. Meanwhile, it is a common practice for buyers to seek help from real estate agents in order to minimize time costs when it comes to housing search.

Baryla and Zumpano (1995) examine buyers search duration considering transactions both with and without real estate agents being involved to see if there is an information asymmetry in the residential real estate market. They consider a cross-section sample of about 500 transactions and apply survival analysis techniques. In particular, they use a parametric method in order to examine buyers search duration and if it is influenced by the fact that intermediaries are involved in a transaction or not. Duration is measured in weeks and its logarithm is used in the analysis. In order to obtain parametric estimations a Weibull distribution of the survival function is assumed and the parameters are estimated by maximum likelihood (ML) method. The data for the analysis is available for the year 1987 through a nationwide survey, as a result, only buyers with complete durations are in the sample, meaning that no censoring problem occurs. The results of the paper indicate that the estimated survival function has a positive duration dependence, meaning that as the duration interval increases the probability that the search will end, gets higher. Also it is found that involvement of real estate agents in the transaction reduces buyers search duration.

Anglin (1997) also looks at the duration of the search time of buyers. He considers two different types of duration. One is measured by the number of houses that a buyer has seen before making a decision. The second one is the classical duration in terms of time, where the search process is measured in weeks spent while looking for a house. Anglin examines the effect real estate agents have on the duration and also if there is a difference in search time depending on the type of house to be bought (resale or newly-built). A sample of 265 buyers is analysed and the data is obtained through a mail survey. No censored data appears as only buyers that have finished the transaction are taken into account. The author tries to answer the question “What is the probability that a buyer will buy a house in the next unit of time (week), given that he/she has not yet brought a house?” by estimating a hazard function. The probability is evaluated using a logit model. Anglin concludes, that people who brought new houses on

³ There are two types of prices considered in the existing literature. The asking price (the price that is stated by a seller in the offer, the starting price) and the actual price the house is being sold for. . The existence of two different prices is due to the bargaining possibilities. In practice it is easier to obtain the data on asking price as the selling price often stays just between the people having a deal.

average searched for a longer period of time than people who bought resold houses. Buyers who brought a house with the help of an agent inspected more houses compared to the ones who were not using an agent.

The buyers' side is also considered in the more recent research, however it is less common to see it as a separate topic of interest. For example, Genesove and Han (2012) apply a random matching model to a survey data on both buyers and sellers. The results they obtain show that both buyers and sellers' TOM decrease with demand. However, the panel data set they use is heavily unbalanced and lacks a number of surveys, which means that the results should be trusted with caution.

In fact, the number of buyers TOM research is relatively small, this is due to the fact that the data for analysis is not easy to obtain. Moreover, if the data is obtained, the surveys are then the most common source, which implies that the data is exposed to a certain level of subjectivity. For this reason it is more attractive for researchers to look at the time-on-the-market from the perspective of house sellers.

2.2. Sellers

Nowadays the duration data from the seller's point of view is often easier to obtain, as it is a common practise to use a multiple listing service when wanting to sell property. So a switch from analysing buyers towards analysing sellers can be seen in recent decades.

One of the early works in this field was by Miller (1978), who analyses the tradeoff between the selling time and price (with respect to both the nominal and real price). A positive relationship is expected between those two variables. A multiple regression is applied to a data set of 91 houses. The results support the initial hypothesis, proving that there is a positive relationship between the TOM and listing price.

Other characteristics affecting the duration except of price also attract attention. For example, Zuehlke (1987) investigates the relationship between the probability of sale and market duration for a sample of 290 single-families (in 1982) out of which only 138 were sold during the considered year. This implies that more than a half of the starting sample contains of censored observations. The author employs a Weibull hazard model for the estimations. He also emphasizes that it is important to take into account a vacancy measure⁴ as it has significant

⁴ Vacancy measure shows if the seller was still living in the house, when one was on sale, or the house was vacant during the marketing period of time.

effect on the final results. Vacant houses are found to exhibit positive duration dependence, meanwhile, for occupied houses almost no duration dependence is found.

Haurin (1988) defines a typical house and then based on a sample of 219 transactions investigates if the fact that a house is substantially different from the typical one affects its marketing time. In his analysis a standard search model is applied, with the assumption of a Weibull distribution being present, even though the problem of censoring is not directly discussed. The conclusion of the paper is that the fact that a house is severely different from the typical house increases the TOM.

It is often the case that there are more than two sides involved in the deal, not only a buyer and a seller, but also a third party representing interests of one of the sides. Donald et al. (1996) investigate the impact brokers and marketing strategies have on TOM in the residential market. A sample of around 2 300 observations is considered with the time horizon of two years which is taken from a multiple listing service. The authors criticize the previously applied approach of regression analysis when it comes to TOM. Instead they use a more appropriate, in their opinion, duration model approach. A survival function is estimated in the analysis, the shape of which resembles the exponential and Weibull distributions. It is found that price-related marketing strategies strongly influence TOM, while no evidence is presented that agents are able to market a house in less amount of time. Also the authors note, that higher prices are associated with higher TOM.

If early research considered data sets were often consisting only from completed durations or the censoring issues did not receive enough attention, the situation changed with time as the research evolved. Genesove and Mayer (1997) consider a right-censored sample, where TOM is measured in weeks. By assuming a proportional hazard rate they obtain the Kaplan-Meier estimator (Kaplan & Meier, 1958) of the survivor function. The estimations suggest that owners with high loan-to-value ratio take longer time to sell their properties compared to owners whose loan-to-value ratio is lower. Glower et al. (1998) consider the role that the seller's motivation plays in determining selling time and both types of price (listing and selling). A standard search model is used for a data set of 115 observations for a bit more than a one year period. Censored observations are present in the sample. A method taking this into account is applied, in particular, a parametric approach assuming a log-logistic distribution and allowing for hazard rate to change over time is used. The results show that a seller, who at the time of listing, has a planned date to move, tends to sell quicker than the one who is not bounded by one.

The relation of duration to price has often been investigated, specifically different types of prices have been considered. For instance, Huang and Palmquist (2001) find a negative impact of the market duration on the reservation price.⁵ The result is obtained through joint estimation method, where a linear system of equations for duration and reservation price is considered. However, in this paper the authors note that there exists potential right censoring in the data, which cannot be treated. Anglin et al. (2003) considers the TOM and degree of over pricing. In the analysed sample the problem of censoring occurs as the sellers can withdraw objects from the market without selling. A parametric approach is applied to deal with the problem. The authors obtain the Kaplan-Meier estimates of the hazard rates and show that they do not vary with duration if not conditioned on explanatory variables. In their analysis they take into account the variation in housing location as well as the seasonal variation. They try to figure out if there is a direct trade-off between the selling price and TOM, but none appears to be present, meaning that an increase in the list price does not cause an increase in duration. Haurin et al. (2010) look at the relationship between the sales and the reservation prices. They also face the problem of censored data and apply proportional hazard models in order to deal with it.

Das (2007) uses a search-theoretical approach to investigate the relationship between probability of sale and market duration in the housing market. A sample of 203 objects is considered, out of which only 103 have completed duration (non-censored). In the analysis a hazard model is employed, which is based upon the reduced-form Weibull model of Lancaster (1979). The obtained results are in line with the ones Zuehlke (1987) had, meaning that positive duration dependence is seen for vacant houses and only little evidence of duration dependence is observed for occupied houses.

It can be seen that during the last decades the amount of research in the field of housing has been increasing. In particular, there are still countries, such as Russia for example, where the real estate market has not yet really been modeled. Furthermore, the methods applied in the empirical analysis are still not perfect, which leaves the opportunity for improvement of estimation techniques when it comes to dealing with censored data.

⁵ Reservation price is the minimum price at which the offer will be accepted by the seller at each point of time.

3. Data

The economic papers related to the duration from a seller's perspective, which were mentioned in the previous section, focus on the United States housing market in different states and cities. The data in those studies were obtained through different multiple listing services. Meanwhile, the current research focuses on the Russian real estate market. Namely, the data for Moscow secondary housing market⁶ is considered.

Information on residential housing is obtained from an online listing service “www.gdeetotdom.ru”.⁷ The sample consists of 10 155 apartments which were on the market in the time period 3rd January 2012 – 1st October 2014. For each object in the data set the date when the apartment was put on the market and when it went off the market is available. Also information on listing prices in roubles is available. In addition, the following characteristics of the property are given: total area of the apartment, floor in the building, location, type of the seller (owner of the property or a real estate agent) and year when the building was built. More detailed information on the variables can be found in the appendix (Table A1).

The main two variables of interest are duration of a house being on the market and its price. The price given in the sample is nominal and for an apartment as a whole. In order to use the price variable for the analysis a number of transformations have been made. First of all, real prices were obtained in order to take inflation into account. Therefore, the price data was corrected by the CPI⁸ for non-food items, taking as basis for correction January 2012. Furthermore, it is important to mention that the prices are given for the apartments, which makes them incomparable between houses of different size. Knowing the number of square meters for each apartment and in order to make the price comparable, the real price per square meter will be used in this paper.

⁶ Secondary housing market includes all the property that is being resold, meaning that all objects on the market are being supplied by the current owners of the property and not by the building company, as no newly-built houses are considered.

⁷ ООО «ГДЕЭТОТДОМ.РУ»

125315, Москва, Ленинградский проспект, д.68, стр.16

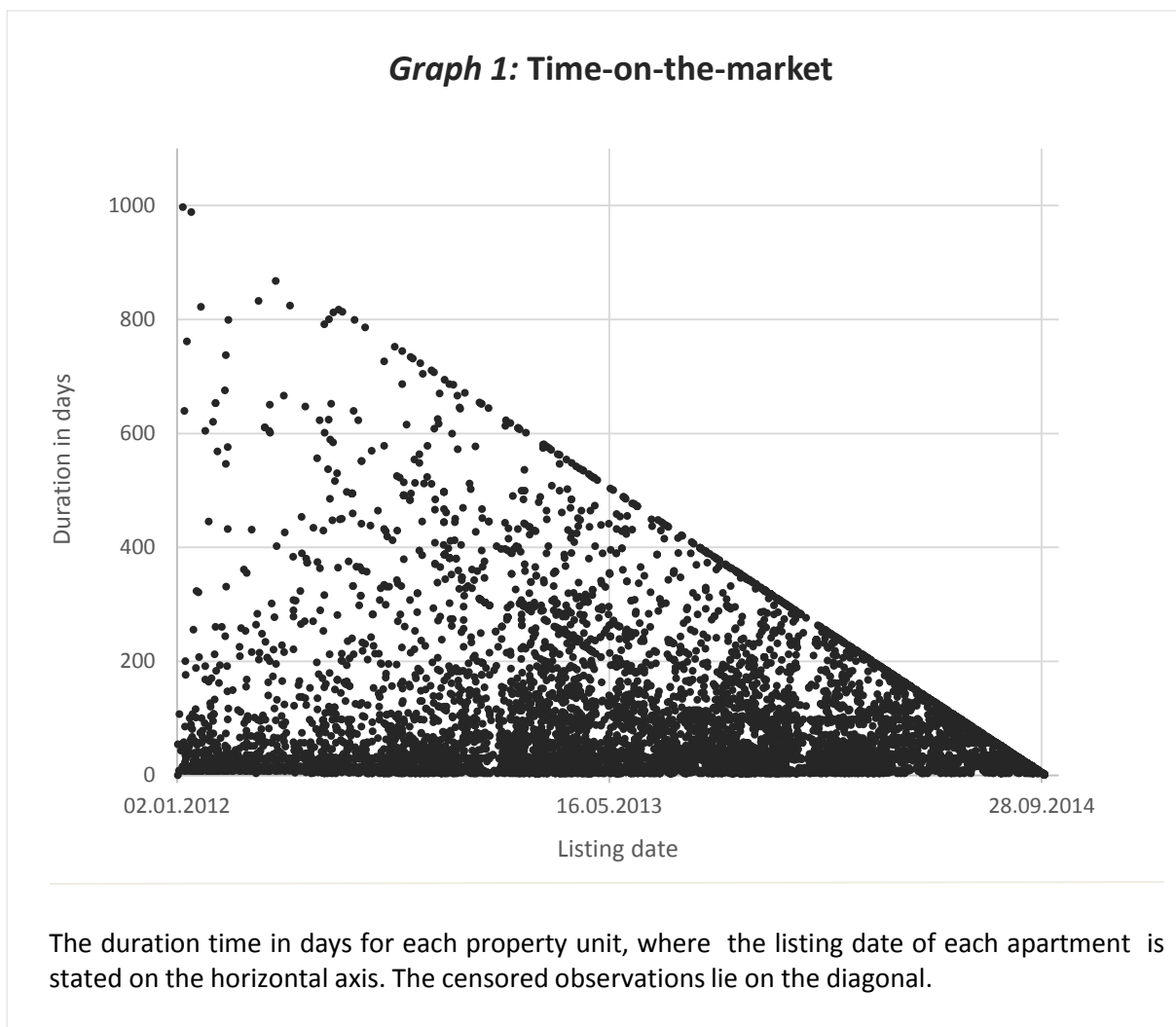
[Limited liability company “gdeetotdom.ru”, Russia, 125315. Moscow, Leningradskiy av. 68/16]

Tel: +7(495)657-99-21

E-mail: info@gdeetotdom.ru

⁸ “Consumer Price Index” for nonfood goods, given in percentage change relative to the previous month. The date is obtained from the Russian Federation Federal State Statistics Service (www.gks.ru).

The other main variable is duration, which reflects for how long the property stayed on the market before being sold. The date when the apartment was advertised on the housing internet portal is the date when the property entered the market, while the date of sale is the one, when the advertisement was taken down from the webpage. This allows calculating the duration as a difference between the time of the sale and the placing time. Both dates are given with high precision, i.e. the exact date and time. However, the duration variable is still not continuous as there is not a house entering the market at every single point of time. Thus, in order to make the analysis more feasible the duration is measured in days.⁹



It is important to pay attention to the problem of right-censoring (subsample of apartments). In particular, the data for analysis was extracted from the web-portal database on 1st October 2014, however at that point of time some of the objects were still on the market.

⁹ To obtain such duration, the values of the duration obtained as the difference of entering and closing dates is rounded up (to the number of days).

This means that for a number of apartments there is no sale date, but only the date when the data was extracted. Such situations imply that for the observations that were still on the market when the data was extracted the TOM is not less than the duration from when the property enters the market until the extraction date. All such observations create a censored subsample and make it impossible to use the duration data for the analysis, as they will necessarily create a bias due to the fact that their duration is underestimated. The problem of censoring can be seen in Graph 1¹⁰ where on the horizontal axis the entry date of the apartment is given and on the vertical – the duration in days. The censored observations are lying precisely on the diagonal. Simply removing the censored observations would remove the diagonal in Graph 1. However, this would not resolve the problem as the remaining observations would be downward biased in duration. This makes it clear that the censored observations need to be treated before the data will be suitable for analysis.

4. Methodology

As previously mentioned, the problem of right-censoring is present in the data. The largest TOM in the considered sample is 997 days and belongs to a completed observation, meanwhile the smallest equals one day and belongs to a censored observation. This implies that the existence of such observations will add a bias to any sort of estimations based on this data. For example, if one wants to estimate the average TOM, the obtained value will have a serious downwards bias, because the observations with a duration less than the actual TOM will be taken into account. Thus, it is necessary to first solve the problem of censored data before moving on to estimating any sort of relationship between TOM and other variables. In order to account for the censoring issue a two-step procedure is suggested. First, a hazard function is estimated. Second, the estimated hazard rates are used in the simulation procedure to obtain the estimates of the duration for the censored observations.

4.1. Estimated hazard function

In order to deal with the censoring problem a nonparametric approach is applied, with the following simulation of the TOM values for the censored observations. The basis for the methodology is formed by the estimation of a discrete-time hazard function.

¹⁰ All graphs, charts and tables presented in the current study are constructed by the author based on the available data and obtained estimations.

To begin with, assume there is a sample of size N available for the analysis.¹¹ Some of the data is right-censored, meaning that the spells are observed from a time 0 until a censoring time t . Both spells with complete and incomplete duration are present in the sample. T_i^* is the true duration, which is only observed for the observations with completed spells. For the censored observations the values of T_i^* are latent. T_i is the observed duration data. It equals completed duration T_i^* if the spell ends before the censoring time and equals to the censored duration¹² C_i^* if the spell has not end when the censoring time occurred. T_i^* and C_i^* are independent from each other (random censoring). It is known which observations belong to the censored subsample, and is indicated by a dummy variable δ_i .

$$T_i = \min(T_i^*, C_i^*), \quad (1)$$

$$\delta_i = \begin{cases} 1, & \text{if } T_i^* < C_i^* \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Furthermore, for the standard survival analysis methods the censoring should be independent, meaning that parameters of the censoring time distribution are not informative about parameters of the completed duration distribution. Then the censoring indicator δ_i may be treated as exogenous (Cameron & Trivedi 2005, pp. 579-580).

Let $t_1 < t_2 < \dots < t_j < \dots < t_k$ ¹³ denote all possible durations (also known as discrete failure times, meaning that the observation fails when the end date appears) occurring in the sample N , ordered from 1 to k , where $k \leq N$. Next, define the following variables:

d_j – the number of observations for which the duration equals to t_j ;

m_j – the number of observations right-censored in the interval $[t_j; t_{j+1})$ ¹⁴;

r_j – the number of observations at risk of failure (the observations that have survived passed the duration length of t_j , i.e. have neither failed nor been censored).

The discrete-time hazard function is defined as the probability of transition (sale) at discrete time t_j given that the observation has survived (has not been sold) until time t_j .

¹¹ The methodological part referring to the nonparametric estimations and survival function is provided according to Cameron and Trivedi (2005). The notations used by the authors are also kept in the current paper.

¹² Same as the censoring time if the starting time is assumed to be 0.

¹³ The time t_j is numerically equal to the duration for the observation existing at this time. For example, $t_1 = 1$, then for all observation that are still existing at t_1 it implies that the observations survived one day and for all of them that failed on this day the duration will be one day.

¹⁴ The number of observations from the censored sub-sample for which duration equals to t_j .

$$\lambda_j = Pr[T = t_j | T \geq t_j] \quad (3)$$

The estimator for this function is then the ratio between the number of observations with duration t_j and the number of observations at risk of failure at the same time t_j .

$$\hat{\lambda}_j = \frac{d_j}{r_j} \quad (4)$$

The discrete-time survivor function is then obtained recursively from the hazard function.

$$S^d(t) = Pr[T \geq t] = \prod_{j|t_j \leq t} (1 - \lambda_j) \quad (5)$$

For the estimation of this function the Kaplan-Meier estimator is used, which is calculated as follows.

$$\hat{S}_t(t) = \prod_{j|t_j \leq t} (1 - \hat{\lambda}_j) = \prod_{j|t_j \leq t} \left(\frac{r_j - d_j}{r_j}\right)^{15} \quad (6)$$

This means that given the duration data it is possible to estimate both the hazard and the survival functions.

The main interest here is the hazard function. After the estimation procedure a series of pairs $(t_j; \hat{\lambda}_j)$ are obtained. Based on the values of $\hat{\lambda}_j$ the complete duration for the censored observations can be simulated. The simulation is carried out for each property unit in the censored subsample in order to obtain the estimation of the completed duration for every element of the subsample. When supplementing the data with the estimates for the censored durations it is then possible to analyse the relationship between the variables of interest.

4.2. Simulation procedures to obtain the parameters of interest

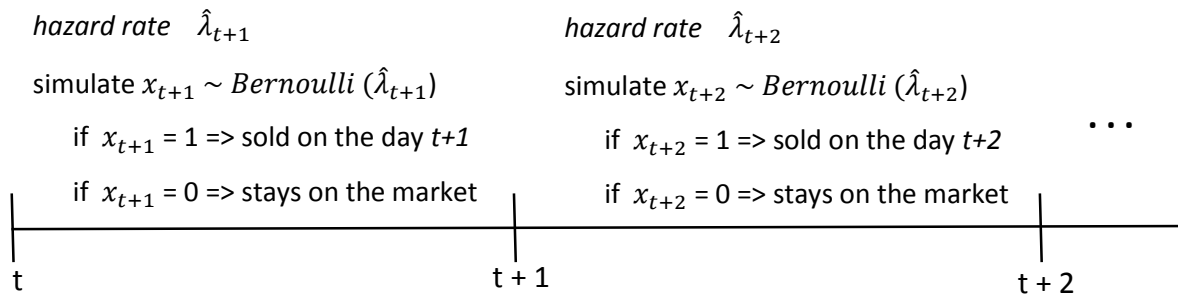
As is explained above, the duration data available for the analysis is right-censored, which makes it unsuitable for empirical analysis. Thus, it is necessary to estimate the completed duration for the censored observations data, because of the bias that is introduced if no adjustments are made. In this section it will be described how such adjustments can be conducted with the help of a simulation technique based on a sequence of Bernoulli distributions.

Start by examining one house from the right-censored subsample. Let this house have some duration of t days, so it is known that this particular house was on the market not less than t days. The information on the days following t is not available, however, the hazard rates

¹⁵ Cameron & Trivedi (2005), pp.578, 581.

have been estimated for all time periods up to t_k . The considered t is less than t_k , because all the censored durations are less than the maximum observed completed durations in the sample. Thus, the hazard rates are used as the parameters of the Bernoulli distributions in the simulation procedure. The Bernoulli distribution is used here as for all the days after the censoring date there is always a binary outcome: the house is either sold or stays on the market.

Chart 1: Duration estimation procedure for the censored data



The second step of estimating the censored duration is based on the estimated hazard function. The time of sale is simulated by drawing from a Bernoulli distribution with the corresponding hazard rate as the parameter until 1 is obtained. This gives the time when property is sold.

For the censored house that has been on the market for t days, consider day $t + 1$ for which a random variable from a Bernoulli distribution with parameter $\hat{\lambda}_{t+1}$ is simulated. This random variable is binary and can take on two possible values: 1 if the house is sold and 0 if it stays on the market. If the value of the simulated variable takes on the value of 1, it means that the house has been sold in the current period implying that the completed duration of the censored observation equals $t + 1$. If, however, the variable equals to 0, then the object stays on the market and the next period should be considered.

The same procedure is then followed, with the only difference that the random variable is now simulated from a Bernoulli distribution with the parameter $\hat{\lambda}_{t+2}$. Then again either the estimated duration indicates a completed duration if the random variable equals to 1, or the procedure continues by considering the next day. The following is repeated until a time $t + m$, where $m \in \mathbb{N}$, has been found such that the random variable takes on the value of 1. With every time period considered, the chance of the observation surviving past the next day becomes less as the survival function is a decreasing function. This implies that eventually there will be

found a time period $t + m$ where each house has left the market. Thus, the estimate of the duration for the censored observation will be obtained.

The described above procedure is illustrated in Chart 1. It is based on creating a sequence of random variables following a Bernoulli distribution. The following procedure is repeated for each of the right-censored observations in the sample, as a result an estimation of the completed duration is provided for each of them. The only problem here that might occur is if the observed duration for a censored observation is close enough to the maximum completed duration appearing in the sample. Thus t is close to t_k . This will lead to a chance of facing a situation when there will be only zeros in the simulated sequence when there is no available hazard rates left. So a completed duration estimation cannot be obtained.

One of the solutions could be to set the actual duration equal to the censored one if no $t + m$ satisfying the condition is found. However, this will unavoidably lead to a downwards bias as it is known that the actual duration is greater than the censored one. In order to avoid this issue the average of the available hazard function values is calculated. Then for all periods $t + m > t_k$ the value of $\hat{\lambda}_{t+m}$ equals the calculated average as long as there is such a time found when the simulated random number from a Bernoulli distribution with the corresponding parameter equals to 1. This means that for some observations of the censored subsample it is possible that the estimated duration is going to be greater than the longest duration observed in the completed duration subsample.

Such an estimation procedure allows obtaining a new series (\hat{T}_i) for the duration variable where no longer the right-censoring problem is present.

$$\hat{T}_i = \begin{cases} T_i^*, & \text{if } T_i^* < C_i^* \\ T_i^{simulated}, & \text{otherwise} \end{cases} \quad (7)$$

It is important to note that if the simulation procedure is repeated several times and different duration estimation series are obtained, then the estimated durations will be different for the same observation in each new simulation. This means that, for example, it is not possible to use this simulation procedure to build forecasts of the duration value for the censored data. It can be seen that such estimation of the censored duration introduces some level of uncertainty in the estimates. This uncertainty will also be present if the relationship between the duration and price is considered.

In order to account for this uncertainty and obtain consistent estimates of the regression parameters the estimated duration series are re-estimated at least 100 times and then the average is taken. This method is based on the idea employed in the bootstrap technique suggested by

Efron (1979). It allows assigning measure of accuracy to sample estimates (Efron & Tibshirani 1993). This helps to partly deal with the estimators uncertainty issue by obtaining consistent estimators after repeating the simulation procedure for a sufficient number of times.

Summing up, in order to obtain the coefficients estimates for the regression between the variables of interest a number of steps should be followed. First, the TOM for the right-censored subsample is simulated and the new duration series (\hat{T}_i) is formed by replacing the duration values for the censored observations by their estimates. Second, the new duration variable is regressed on the price and the coefficient estimates are obtained via ordinary least squares (OLS). Then following the logic of the bootstrap technique this procedure is repeated a large number of times (L), which gives L different values of the coefficient estimates. Lastly, the average of the obtained estimates is taken, which exactly gives the regression coefficient estimator.

The described above method is effective in dealing with the right-censoring issue in the data when wanting to estimate the relationship between a censored variable and other variables of interest. The fact that it is based on a simulation technique helps to avoid making extra assumptions. In particular, no assumptions about the distribution of the data are made, thus the method belongs to the category of non-parametric estimation techniques.

5. Empirical Analysis

Before applying the considered earlier methodology to the available sample, it is first necessary to emphasize the following. The data set consists of 10 155 observation, which implies $N = 10\ 155$. Out of those N observations 1 229 belong to the censored subsample. For each of the right-censored observation a simulation is carried out to obtain the estimated TOM

The following section consists of the general analysis of available data and correction for factors which influence the price making it incomparable. Then the hazard and the survival functions are estimated. Based on those estimates a simulation procedure is carried out to provide the estimates of completed duration for the censored observations. When both price and duration are adjusted, a regression for those variables is estimated.

5.1. Preliminary analysis

The price variable considered so far is the real price of property per square meter. Thus, the floor space of the apartment is taken into account. Also a correction for inflation has been

made based on the CPI for the non-food items, the details on the correction can be found in the data section. Even though inflation was extracted from the data, the price range is relatively large. This means that there is a chance that other property characteristics affecting the price should be taken into account. Except of the floor space, the location of the apartment also should influence the price.

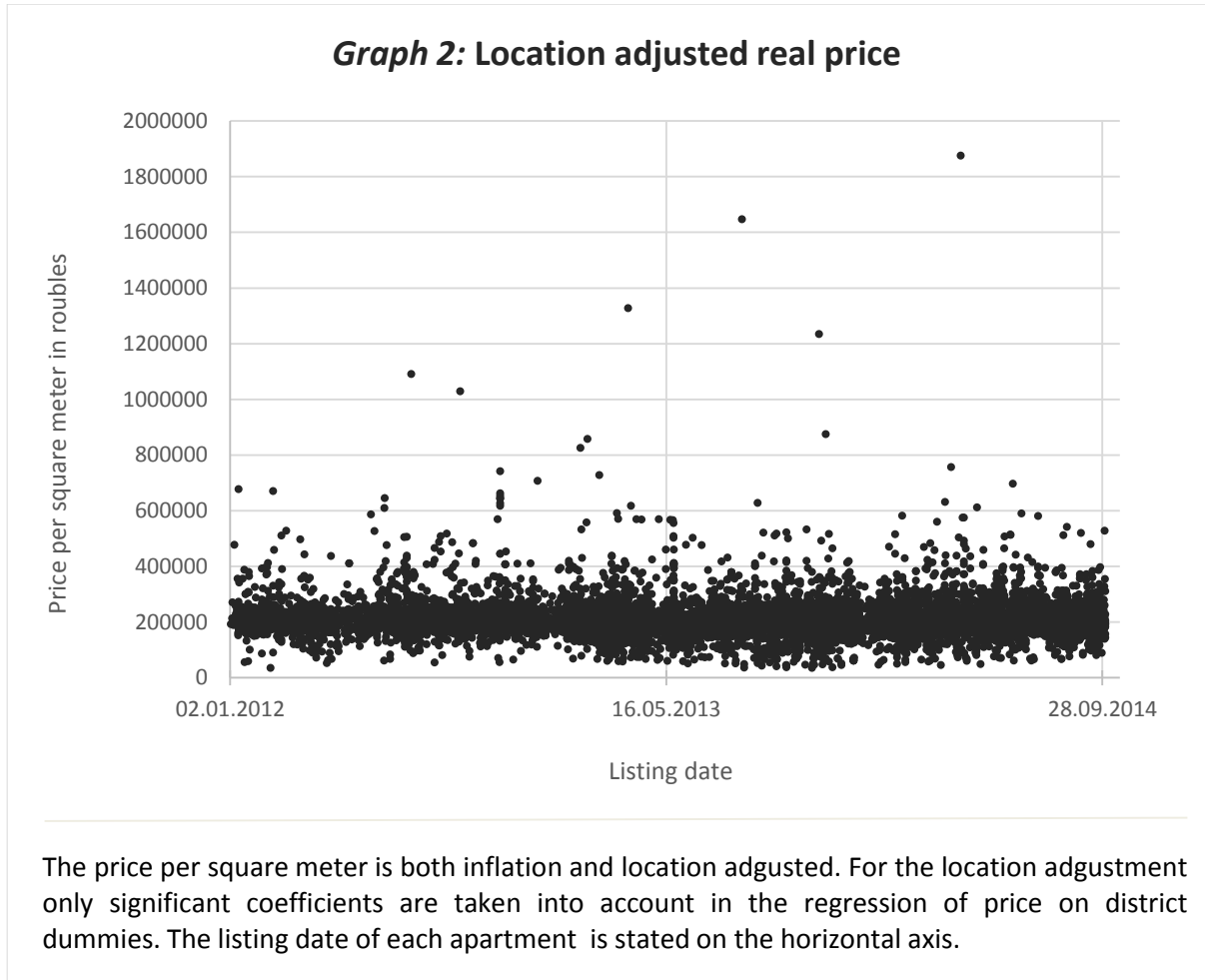
In the data set the address of each property unit is given, however this exact address of each unit would be unnecessarily overcomplicating the analysis. Thus, it is convenient to consider specific location grouping in order to make the analysis feasible. There are different administratively defined territorial divisions of Moscow depending on the size. It seems the most reasonable to consider districts¹⁶ as their size is relatively small, although the number of them in the city still makes it possible to carry out the analysis. There are 125 districts in Moscow and property objects from each of them are present in the sample. In the data set the variable SUB_LOCALITY is the one responsible for the districts. In order to examine the effect of the area on the price 125 dummy variables are created (one per each district) and inflation corrected prices are regressed on the district dummies.¹⁷ OLS is used to obtain the estimates. When looking at the p-values most of the coefficients are significant at the 1% level.¹⁸ This implies that the location of an apartment has an influence on its price.

Thus inflation adjusted price per square meter is also corrected with respect to the location. All significant coefficients are taken and subtracted from the price of the corresponding property units. After taking the district parameter into account, the average real price for a square meter in the considered sample is 211 870 roubles. The location corrected prices are presented in Graph 2. It can be seen that several of the observations have price values which lie outside of the main range. This could be due to the fact that most of the property objects are sold directly by the owners, which means there is a probability of them overpricing their property in the current market conditions. In this case, the owners who are not well informed about the market may expect to receive a higher income, than the actual market price of the property object.

¹⁶ rus. – район

¹⁷ In order to avoid multicollinearity only 124 dummies are present in the regression, because the estimations are carried out for a linear model with a constant.

¹⁸ Due to a larger number of variables the results of estimations are not stated in this paper. The regression estimates can be provided by the author upon request.

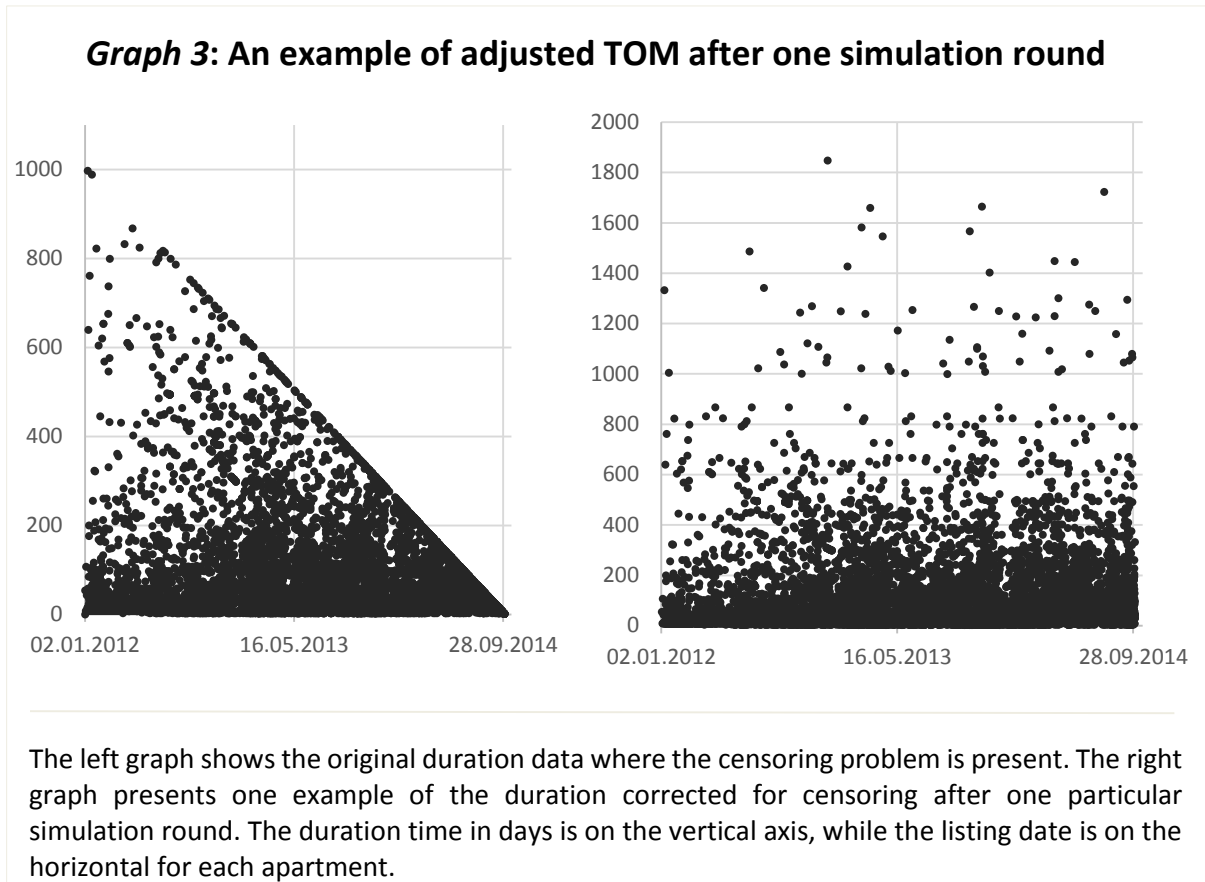


5.2. Duration simulations

The completed duration part of the data set consists of the TOM, which vary from 1 up to 997 days, implying that t varies in this interval. Then for each t in the interval the following parameters are estimated: d_t , m_t , r_t ; and then the hazard function estimates are obtained: $\hat{\lambda}_t$, $t = 1, \dots, 997$. For values of t greater than 997 the average value of the hazard rate is considered, which equals to 0.005. The survival function is also obtained using the Kaplan-Meier estimator (see Appendix A2). The estimated function is decreasing, which goes in line with the theory, meaning that the probability of surviving one more day is less with every next day coming. Therefore, positive duration dependence is found.

As can be seen from Graph 3 the presence of right-censoring does not allow to use the data for any further analysis without preliminary correction. Thus, based on the hazard function, estimates of the duration time for the observations from the right-censored subsample are

simulated. The procedure described in the methodology section is applied.¹⁹ The new TOM series corrected for censoring are presented in Graph 3.



As mentioned before, it is not possible to use just one series with estimated duration for censored observations to draw conclusions about the characteristics of the duration or its relationship with other variables. Thus, the simulation procedure is carried out L times, which gives L possible realisations of the TOM. By calculating the average of the series averages it can be then said, that the average TOM for a property unit is around 104 days, which is approximately 3.5 months. This result goes in line with other research. For instance, Zuehlke (1987) shows that in the beginning of 1980s the average market duration was about 4 months when considering Tallahassee, Florida housing market. Haurin (1988) finds that a typical TOM for resale residential housing in Columbus, Ohio was 2-3 months during the late 1970s. Donald et al. (1996) show that mean TOM was 108 days in North Carolina in the beginning of 1990s. Das (2007) provides information on the average duration for a non-censored part of the data from New Orleans, which equals 3 months. If considering the completed observations

¹⁹ All the simulations and estimations are carried out in MATLAB. The code for the simulation procedure and other estimations is written by the author and can be provided upon request.

subsample of the currently analysed data set, then the average market duration is about 2.5 months. It is interesting to notice, that even though the focus of the current paper is on the Moscow resale real estate market and not on the United States, the average time-on-the-market is relatively similar. This provides evidence that the real estate market seems relatively stable both with time (as data for various years was considered by different researchers) and place as common characteristics are found. Subsequently, it is of interest to see if the same relationship between market duration and other variables, such as price, holds for the Moscow property market as for the U.S. markets.

5.3. Estimation results

Before estimating the relationship between the TOM and housing price, it is necessary to state that the logarithms will be considered. As the value range of both variables is relatively large such scaling makes the analysis more feasible. The graphs for the variable logarithms can be found in the Appendix (A3).

$$l_{pm} = \ln(\text{location adjusted real price per } m^2)$$

$$l_{dur} = \ln(\text{duration adjusted for censoring})$$

To begin with the market duration is regressed on the price. The coefficients of the model described by equation (8) are estimated. OLS is applied to calculate the parameter values.

$$l_{dur} = \beta_0 + \beta_1 l_{pm} + \varepsilon \quad (8)$$

Estimates for this regression are calculated for each of the corrected duration series obtained after the simulation process. Then an average of all L coefficients is found, which gives a consistent estimate of the parameter of interest. The same procedure is applied in order to obtain the p-value for the regression coefficient, where a 1% significance level is taken into consideration. The results of the estimations are presented in Table 1. The coefficient can be interpreted as the cost of a price change in terms of the amount of extra time that the apartment stays on the market if the price is increased by some time. Thus, it can be seen that a 1% increase in the real price will induce a 0.46% increase in market duration at any reasonable level of significance. This means that for the average house price of 211 870 roubles and an average duration of 104 days, an increase of the price by approximately 2 119 roubles will lead to an increase of the TOM by almost 5 days. This result goes in line with the ones obtained by Miller (1978) and Donald et al. (1996), specifically, that there is positive relationship between the house price and its TOM.

Table 1: Regression estimations

<i>Dependent variable:</i>	<i>l_dur</i>	
	(1)	(2)
l_pm	0.46 (6.1e-23)	0.43 (5.6e-21)
USER_TYPE		-0.52 (1.2e-77)

The results of the OLS estimations for the regressions of interest. The presented coefficient and p-values are the average ones, obtained from L estimated regressions for each of the duration corrected series.

As described in the literature review several researchers state that the presence of a third party in the transaction has a significant effect on the duration. In the available data set information on type of the seller is present. The house is either sold by the owner directly or with the help of a real estate agent. Then it is possible to include the variable in the specification of the regression in order to see if the presence of real estate agents affects the TOM on the Moscow housing market. The variable responsible for the type of a seller is a dummy, which takes on the value of 1 if the property is being sold directly by the owner. Including the variable into the model the following specification is estimated.

$$l_dur = \beta_0 + \beta_1 l_pm + \beta_2 USER_TYPE + \varepsilon \quad (9)$$

The values of the estimated coefficient can be found in Table 1. Both of them are significant at the 1% level. The relationship between the price and duration stays positive and the change in the coefficient value is minor. When the type of the seller is taken into account a 1% increase in price is accompanied by a 0.43% increase of the TOM. Therefore, for an average apartment, an increase of the price by 2 119 roubles increases the market duration by about 4.5 days. It is also proved that the duration depends on who was selling the property. In particular, it takes 0.52% more time for the sellers who sought help from a realtor to achieve a deal with the buyer. So on average the owner spends 0.5 day less on selling if not using a real estate agent. This might be due to the fact that sellers who go to real estate agents are the ones who are less familiar with the market situation, thus, have higher expectations, which makes it harder for them to agree on a deal with a buyer. Another possible explanation could be that the sellers hiring an agent from the start have fixed conditions under which they want to perform the

transaction, therefore, it takes a longer time to find a buyer who will agree to those conditions. The obtained result of negative effect that the real estate agent has on the market duration goes in line with conclusions made by Anglin (1997).

Summing up, a positive relationship between the price and duration has been found for the Moscow real estate market, which means that the same tendencies hold for it as for the United States housing market. Thus, the housing market seems to exhibit similar characteristics regardless of time and country specifics, as the results go in line with the ones obtained for various housing markets in various time periods. However, this conclusion holds only when considering markets with similar structure, meaning that the selling process functions in the same manner and the property rights are of similar structure.

6. Conclusions

Even though the housing market has a number of similarities with other markets, it also has characteristics that make it very different to examine and in some sense also make it unique. The objects of sale on this market are durable goods, implying that both buyers and sellers are ready to spend some time in order to find the most suitable deal for them. Thus, the time-on-the-market is one of the most important parameters in the analysis of the real estate market.

Duration, as the time between the listing and selling dates, has been of interest to researchers for a long time. However, the problem of censoring in the data is a serious obstacle for empirical analysis. Most of the research in the field of interest has been focusing on the parametric approach of dealing with right-censoring issues. The main disadvantage of such a method is, however, that it requires strong distributional assumptions and does not perform well if the distributional assumption fails to hold. In the current study a nonparametric approach is considered instead. It is employed to estimate the hazard rate and the survivor function. Based on the estimated survivor function the conclusion is drawn that the data exhibit positive duration dependence. The estimated hazard rate is employed to simulate the duration for the observations belonging to the right-censored subsample. Using the corrected duration data the regression coefficients for the models of interest are obtained.

When analysing the Moscow secondary housing market, the real price per square meter turns out to be dependent on the location. Therefore, it has been corrected for it before proceeding with the analysis. A significant positive relationship is found between the price and time-on-the-market, implying that the increase of the property price comes with the cost of increasing time until the property gets sold. Furthermore, it has been found that on average the

owners who are selling their property directly and do not rely on help of real estate agents, tend to sell their property quicker. Thus, it has been empirically proved that for the Moscow real estate market both the price and type of the seller affect the market duration. It could also be possible that there are other factors influencing the duration on the housing market, which were not represented in the available data. This could be a topic for a further investigation.

7. References

Anglin, P.M., 1997. Determinants of buyer search in a housing market. *Real Estate Economics*, 25, pp.567–589.

Anglin, P.M., Rutherford, R. & Springer, T.M., 2003. The Trade-off between the Selling Price of Residential Properties and Time-on-the-Market: The Impact of Price Setting. *Journal of Real Estate Finance and Economics*, 26(1), pp.95–111.

Baryla, E.A. & Zumpano, L. V., 1995. Buyer Search Duration in the Residential Real Estate Market: The Role of the Real Estate Agent. *Journal of Real Estate Research*, 10(1), pp.1–14.

Benefield, J.D., Cain, C.L. & Johnson, K.H., 2011. On the Relationship Between Property Price, Time-on-Market, and Photo Depictions in a Multiple Listing Service. *Journal of Real Estate Finance and Economics*, 43, pp.401–422.

Cameron, A.C. & Trivedi, P.K., 2005. *Microeconometrics: Methods and Applications*, Cambridge University Press.

Chay, K.Y. & Greenstone, M., 2005. Does Air Quality Matter? Evidence from the Housing Market. *Journal of Political Economy*, 113(2), pp.376–424.

Das, A., 2007. Testing a hazard model for the housing market in New Orleans. *American Journal of Economics and Sociology*, 66(2), pp.443–456.

Díaz, A. & Jerez, B., 2013. House prices, sales, and time on the market: A search-theoretic framework. *International Economic Review*, 54(3), pp.837–872.

Donald, J., Terry, S. & Daniel, W., 1996. Time on the market: the impact of residential brokerage. *Journal of Real Estate Research*, 12, pp.447–458.

Efron, B., 1979. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1), pp.1–26.

Efron, B. & Tibshirani, R.J., 1993. *An Introduction to the Bootstrap* D. R. Cox et al., eds., Chapman & Hall.

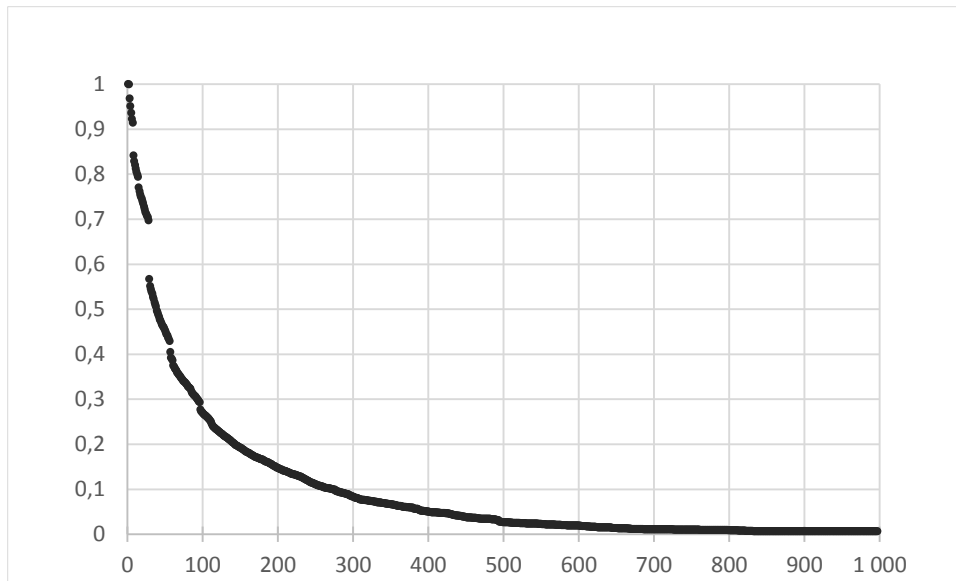
- Genesove, D. & Han, L., 2012. Search and matching in the housing market. *Journal of Urban Economics*, 72(1), pp.31–45.
- Genesove, D. & Mayer, C., 1997. Equity and time to sale in the real estate market. *The American Economic Review*, 87(3), pp.255–269.
- Glomer, M., Haurin, D.R. & Hendershott, P.H., 1998. Selling time and selling price: The influence of seller motivation. *Real estate economics*, 26(4), pp.719–740.
- Haurin, D.R. et al., 2010. List Prices, Sale Prices and Marketing Time: An Application to U.S. Housing Markets. *Real Estate Economics*, 38(4), pp.659–685.
- Haurin, D.R., 1988. The Duration of Marketing Time of Residential Housing. *Journal of the American Real Estate & Urban Economics Association*, 16(4), pp.396–410.
- Huang, J.C. & Palmquist, R.B., 2001. Environmental Conditions, Reservation Prices, and Time on the Market for Housing. *Journal of Real Estate Finance and Economics*, 22(2/3), pp.203–219.
- Jaén-García, M. & Piedra-Muñoz, L., 2012. An empirical model for housing tenure choice and demand in the Spanish market. *Journal of Post Keynesian Economics*, 34(4), pp.653–684.
- Kaplan, E.L. & Meier, P., 1958. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 53(282), pp.457–481.
- Lancaster, T., 1979. Econometric Methods for the Duration of Unemployment. *Econometrica*, 47(4), pp.939–956.
- Miller, N.G., 1978. Time on the Market and Selling Price. *Journal of the American Real Estate and Urban Economics Association*, 6(2), pp.164–174.
- Smith, L.B., Rosen, K.T. & Fallis, G., 1988. Recent Economic Developments in Models of Housing Markets. *Journal of Economic Literature*, 26(1), pp.29–64.
- Stigler, G.J., 1961. The Economics of Information. *Journal of Political Economy*, 69(3), p.213.
- De Wit, E.R., Englund, P. & Francke, M.K., 2012. Price and transaction volume in the Dutch housing market. *Regional Science and Urban Economics*, i, pp.1–54.
- Zuehlke, T.W., 1987. Duration dependence in the housing market. *The Review of Economics and Statistics*, 69(4), pp.701–709.

8. Appendix

8.1. A1. Table: Description of the Variables

Variables	Name	Unit of measure	Comments
<i>Advertisement number</i>	ID_ADVERT	–	Unique for each property unit.
<i>Property listing date</i>	DT_CREATE	date	The day when the advertisement was posted on the portal.
<i>Property selling date</i>	DT_END	date	The day when the advertisement was taken down from the portal.
<i>Type of the seller</i>	USER_TYPE	binary	1 if the owner was selling the property on his own, 0 if a real estate agent was involved in the process.
<i>Floor number</i>	FLOOR_QTY	–	The number of the floor where the apartment is located in the building.
<i>Number of rooms</i>	ROOM_QTY	–	
<i>Property area</i>	FULLSQUARE	m^2	Total area of the apartment.
<i>Living area</i>	LIVESQUARE	m^2	The area excluding kitchen and bathroom.
<i>Kitchen area</i>	KITCHENSQUARE	m^2	
<i>Price</i>	PRICE	roubles	Given prices are nominal, they are the ones seller states when listing the property.
<i>Address</i>	ADDRESS	–	
<i>Moscow districts (район)</i>	SUB_LOCALITY	–	Apartments from 125 different districts are present.
<i>Nearest subway station</i>	METRO	–	In some cases more than one station per property is given.
<i>Time to the nearest subway station</i>	Minut do metro	–	less than 5 minutes 5-10 minutes 10-15 minutes 15-20 minutes 20-30 minutes 30 minutes – 1 hour more than 1 hour
<i>Type of transport</i>	Transport to metro	binary	1 – by public transport 0 – by foot

8.2. A2. Graph: Kaplan-Meier estimator of the survival function



8.3. A3. Graph: Logarithms of market duration and price

