# CHARLES UNIVERSITY IN PRAGUE

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies


## Bachelor thesis


**2015** **Tereza Sklenářová**

# CHARLES UNIVERSITY IN PRAGUE

## FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies



**Tereza Sklenářová**


# Empirical Analysis of Prague Flat Market

*Bachelor thesis*


Prague 2015

**Author**: Tereza Sklenářová

**Supervisor**: Mgr. Tomáš Křehlík M.A.

**Academic Year**: 2014/2015

## Bibliographic note

SKLENÁŘOVÁ, Tereza. *Empirical Analysis of Prague Flat Market*. Prague, 2015. 35 p. Bachelor thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies. Supervisor: Mgr. Tomáš Křehlík M.A.

## Abstract

The purpose of this work is to model the prices of real estate, concretely of Prague flats, which belong to the most important economic indicators. In the theoretical part, the main housing market participants are defined, special features of housing markets are described and most frequently used valuation methods are discussed. Most attention is focused on so called hedonic pricing model, which is applied as a base for the pricing equation in the econometric part. This is carried on various subsets of public available data regarding the characteristics of Prague flats, using ordinary least squares as well as weighted least squares. Several hypotheses about the relationship between the price and the explanatory variables are tested before creating the final model. The results are commented and compared with literature concerned with the same topic in other locations.

## Abstrakt

Cílem této práce je modelovat ceny nemovitostí, konkrétně pražských bytů, které patří k nejdůležitějším ekonomickým indikátorům. V teoretické části jsou definováni hlavní účastníci trhu s bydlením, popsány speciální rysy trhu s bydlením a diskutovány nejčastěji používané metody oceňování. Největší pozornost je zaměřena na takzvaný hedonický cenový model, který je použit jako základ pro cenovou rovnici v ekonometrické části. Ta je prováděna na různých podmnožinách veřejně dostupných dat týkajících se charakteristik pražských bytů, používá metodu nejmenších čtverců a metodu vážených čtverců. Před vytvořením závěrečného modelu je testováno několik hypotéz o vztahu mezi cenou a vysvětlujícími proměnnými. Výsledky jsou komentovány a porovnány s literaturou zabývající se stejným tématem v jiných lokalitách.

## Keywords

Housing market, real estate appraisal, hedonic regression, Prague, econometrics, ordinary least squares, weighted least squares

## Klíčová slova

Trh s nemovitostmi, oceňování nemovitostí, hedonická regrese, Praha, ekonometrie, metoda nejmenších čtverců, metoda vážených čtverců

**Range of thesis:** 72.073 symbols (including spaces)

## Declaration of Authorship

1. The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

2. The author hereby declares that all the sources and literature used have been properly cited.

3. The author hereby declares that the thesis has not been used to obtain a different or the same degree.

Prague 31.7.2015 Tereza Sklenářová

## Acknowledgments

I would like to thank the supervisor of my bachelor thesis, Mgr. Tomáš Křehlík M.A., for providing me with the econometric data.

# Institute of Economic Studies
# Bachelor thesis proposal

**PROPOSED TOPIC**

Empirical Analysis of Prague Flat Market

**TOPIC CHARAKTERISTICS**

The purpose of this work is to model the prices of real estate, which are an important indicator in economics, using microeconomic data and different econometric methods. Primarily, I want to concentrate on real estate appraisal. I will focus on providing robust results testing specifications using various methods and different specifications. The results will be compared with literature concerned with the same topic in other locations.

**HYPOTHESES**

I would like to concentrate mainly on the following hypotheses:

* Is the relationship between the distance from the city centre and the price quadratic?
* Is the relationship between the square footage and the price exponential?
* How does the floor affect the price?
* Do people prioritize the kitchen or the kitchenette?
* Does the square footage have the largest impact on the price?

**METHODOLOGY**

In the pricing part of the thesis I will use hedonic microeconomic models as a base for the pricing equation. The econometric part will be carried on various subsets of data using ordinary least squares, robust variants of linear models (generalised least squares or iteratively reweighted least squares).

**OUTLINE**

1. Introduction
2. Literature review
    a. Housing market
    b. Hedonic models in microeconomics
3. Methodology
4. Results

a. Data description

b. Econometric results

5. Conclusion

**BIBLIOGRAPHY**

MALPEZZI, S. (2002): "Hedonic Pricing Models: A Selective and Applied Review." *Section in Housing Economics and Public Policy: Essays in Honour of Duncan Maclennan*

MILES, D. D. (1994): *Housing, financial markets and the wider economy*. Chichester: Wiley. ISBN: 0471952109

MONSON, M. (2009): "Valuation Using Hedonic Pricing Model." *Cornell Real Estate Review* 7: pp. 62-73

SHEPPARD, S. CH. (1998): "Hedonic Analysis of Housing Markets." *Handbook of regional and urban economics* 3: pp. 1595-1635

SIRMANS, G. S., D. A. MACPHERSON & E. N. ZIETZ (2005): "The Composition of Hedonic Pricing Models." *Journal of Real Estate Literature* 13.1: pp. 1-44

# Contents

# Introduction

The aim of this thesis is to model the prices of real estate (concretely of Prague flats), which are, along with inflation, interest rates and stock indexes, one of the most important economic indicators. As most of the US post-war recessions had roots in real estate markets and in credit inflation connected with the boom on this market, it is clear why it makes sense to follow up the real estate market more carefully than the other markets.

The 2008 financial crisis, which was a direct consequence of US mortgage crisis, is believed by many economists to have been the worst financial crisis since the Great Depression of the 1930s. It soon expanded to Ireland, Great Britain, Spain and Baltic States, where the price growth from the previous years was not sustainable as well. In some countries, the real estate prices fell by almost 30 % and, as a result, their reputation of stable and profitable investment was disrupted.

According to Nakamura (2010), one source of these mortgage problems has been the validity of the home appraisals. As he shows in his article, the appraisals have been biased upward, which made the mortgages riskier. Many house prices were namely below the face value of mortgages and, as a result, the expected return on many mortgages has tumbled since the collateral, one of major forces supporting mortgages, has weakened. That is one of the reasons why correct real estate appraisal, which is the main part of my thesis, is crucial.

The thesis is divided into three main parts. The first part deals with the housing market. Its most important participants are defined and the special features of housing market, which distinguish it from standard commodity markets, are described. Most attention is focused on the real estate appraisal – the most frequently used valuation methods are discussed in detail.

The second part is concentrated on so called hedonic pricing model, which is an alternative valuation method that can be used even when the other methods described in the first chapter fail. Hedonic pricing method tries to decompose the house into its individual characteristics such as size, number of rooms or distance from the city centre and, subsequently, determine the effect of each of those characteristics on the price. In the second chapter, this methodology is described, including the discussion of its advantages and disadvantages. Its brief history and list of the most often used housing characteristics are also part of this chapter.

In the last, most important chapter, the hedonic pricing method is applied in practice. Based on the public available data regarding the characteristics of Prague flats, I will be able, besides other things, to answer the following questions:

- Is the relationship between the distance from the city centre and the price quadratic?
- Is the relationship between the square footage and the price exponential?
- How does the floor affect the price?
- Do people prioritize the kitchen or the kitchenette?
- Does the square footage have the largest impact on the price?

# 1 Housing market

InvestorWords online (2015) defines housing market as the market of houses being purchased and sold between buyers and sellers either directly or indirectly through brokers. In this chapter, its main participants are defined and its special features are described. However, most attention is focused on real estate valuation methods.

## *1.1 Housing market participants*

Housing market, as well as any other markets, is formed by the demand side and the supply side. Firstly, it is useful to define the main participants in both sides of this market. The following list is based on Hopkins and Acton (1999):

- **Owners/users** – people who purchase house or commercial property[1] as an investment as well as to live in or utilize as a business. They are also both owners and tenants;
- **Owners** – these people do not consume the real estate they purchase, but rather rent out or lease it to someone else. They are pure investors;
- **Renters** – people who pay rent for the use of the owner’s property. They are pure consumers;
- **Developers** – these people develop real estate, especially by preparing a site for residential or commercial use, which results in new product for the market;
- **Renovators** – people who are suppliers of renovated buildings to the market.

The owner/user, owner and renter constitute the demand side of the housing market, the other form the supply side. Last but not least participants of the market are so called **facilitators**, whose task is to facilitate both purchase and sale of real estate. They include for example banks, lawyers and real estate brokers (Hopkins and Acton, 1999).

## *1.2 Special features of housing market*

Housing market, compared to standard commodity markets, has quite a lot of special features, which are important to bear in mind whenever we want to analyse it. The most important characteristics are described here, based on Hilbers et al. (2008), Lux (2009) and Miles (1994):

- **Durability** – thanks to modern technologies, the useful life of houses is getting longer, which means that houses with substantially different ages can exist at the same time in the same market. According to the research of Raiffeisen stavební spořitelna (2013), in 2013, two thirds of the Czech population lived in houses or flats older than 30 years and 7 % even in the dwellings which were built before more than 100 years. The situation is similar in most European cities (Miles, 1994). As a result, housing is a subject to both consumption and investment (Lux, 2009);
- **Heterogeneity** (or uniqueness) – this means that every unit of real estate is unique in terms of its characteristics (e.g. age, floor space, furnishings, location or quality of the environment), which is not the case of other consumer durables such as computers, cars or tools;
- **Inelasticity of supply** – this feature is a direct consequence of heterogeneity. In the first place, it is not possible to purchase a particular dwelling if its owner does not want to sell it. Secondly, in case of new houses, the supply is restricted (and therefore often lags behind demand impulses), especially because of planning regulations, limited availability of skilled labour, time delays in construction projects and so on. Moreover, large increases in house prices obviously do not automatically lead to significant increases in the amount of land which is available for housing developments. If the demand suddenly slows down, the supply response will be lagged as well (Hilbers et al., 2008). As a result, in the short run, there is a great potential for disequilibrium in the housing market and at the same time, the adjustment mechanism is relatively slow, compared to more fluid markets;
- **Collateral** – the households are able to raise loans against housing collateral. In many developed countries, it is possible for people to borrow money, required to finance house purchase, under more favourable conditions (i.e. more money at lower cost) than for other purposes. In the Czech Republic, mortgages formed the major part of the debt of the personal sector (circa 65 %) in 2014 (Czech National Bank, 2014);

[1] Real estate property that is used for business activities (e.g. industrial properties, shopping centres and offices)

- **Well-developed secondary market** – this is one of the reasons why housing is a good collateral. However, the turnover on this market is relatively low due to substantial transaction costs associated with house purchase;
- **High transaction costs** – they include for example land transfer taxes, legal fees, moving costs, real estate fees and search costs, but also time and emotions invested. In some European countries, the transaction costs reach up to 15 % - 20 % for both buyer and seller, especially if the intermediaries such as real estate agencies are included (CMS Legal Services EEIG, 2011). Therefore they can significantly influence the selection of future dwelling itself;
- **Price volatility** – although for instance exchange rates, bond prices and equity values may be much more volatile, the implications of changing house prices on the distribution of wealth are often even more important. If the proportion of owner-occupancy[2] is high, the rising real house prices will cause more equal ownership of wealth. On the other hand, if it is low, the inequality is expected to grow (Miles, 1994);
- **Tax treatment** – the tax regime tends to favour home ownership in many countries, which strongly affects conditions in housing markets. They can be influenced for example by real estate taxes, the tax deductibility of certain costs (e.g. mortgage interest payments) and housing subsidies (Hilbers et al., 2008);
- **Financial intermediaries** – finance companies such as commercial banks, mortgage banks or building societies are usually deeper involved in the housing market than in the markets for other consumer durables. According to the research of *Equa bank*, in the Czech Republic, 20 % of people between 25-30 years were already home-owners in 2013 (Studentskéfinance.cz, 2013). Thus, if the cost of houses and the early age at which people buy them are taken into account, it is not surprising that lending practices of financial firms have a great impact on the demand for housing;
- **Immobility** – real estate is spatially fixed[3], which implies that there cannot be a physical marketplace. Because of this spatial fixity, people have to adjust the market by moving to dwelling units, rather than moving the dwellings. Nevertheless, it is important to bear in mind that purchase of a particular dwelling includes also purchase of the socio-economic status of a

[2] A form of housing tenure where a person, called the owner-occupier, owns the home in which they live

neighbourhood and the level of accessibility to the place of employment, which affects the price as well (Lux, 2009);

- **Low liquidity** – compared with financial assets, housing is a relatively illiquid good. Furthermore, it is a very expensive good, so people tend to be extremely careful before making the final decision to buy it and it usually takes them more time than in case of other consumer goods;
- **Imperfect knowledge** – for both the buyer and the seller, it is not possible to acquire perfect knowledge about the situation on highly dispersed housing market. To behave rationally, it would be namely necessary not only to know the price of a dwelling, but also the prices of particular housing attributes in diverse regions and locations;
- **Varying conditions of sales** – part of bilateral negotiations about price of a dwelling consist also of negotiations about the condition of the property (”as is” as opposed to after renovation or certain repairs) and other aspects of the sale, e.g. distribution of costs or timing (Hilbers et al. 2008);
- **Varying financing conditions** – this mainly concerns the presence of specialized mortgage finance institutions and mortgage-backed securities markets, options to refinance and the use of real estate as collateral, and the supervisory and regulatory framework for housing finance (Hilbers et al., 2008).

## 1.3 Real estate appraisal

Real estate appraisal is defined by Investopedia online (2015) as its valuation by the estimate of an authorized person. The goal is usually to find so called market value, whose common definition was set by International Valuation Standards Committee as follows: ,,*Market value is a representation of value in exchange, or the amount a property would bring if offered for sale in the open market at the date of valuation under circumstances that meet the requirements of the market value definition.*“ For the purpose of standards, market value is understood as ,, t*he estimated amount for which an asset should exchange on the date of valuation between a willing buyer and a willing seller in an arm's length transaction*[4]*, after proper marketing, wherein the parties had*

[3] Locationally immobile

[4] A transaction in which the buyers and sellers of a product act independently and have no relationship to each other

*each acted knowledgeably, prudently and without compulsion" (*Australian Valuations, *2015).*

In case of real estate transactions, appraisals are frequently required, especially because of the uniqueness of each property (see the previous chapter). Moreover, compared to e.g. corporate stocks, which are traded on a daily basis, real estate transactions occur much less often. Valuations of real estate are used not only in case of purchase or sale, but also for transfer, tax assessment, expropriation, inheritance or estate settlement, investment and financing (Pagourtzi et al., 2003).

Valuation of real estate should provide a quantitative measure of the benefits and liabilities that result from the ownership of the real estate. It can be carried out by various players in the marketplace, e.g. real estate agents, appraisers[5], assessors[6], mortgage lenders, brokers, property developers, investors and fund managements, lenders, market researches and analysts, shopping centre owners and operators or other specialists and consultants (Pagourtzki et al., 2003).

There are many methods of valuation in the world. Most of them rely upon some form of comparison, including direct capital comparison, which is the simplest form, or range of observations that enable the valuer to determine a regression model. Pagourtzki (2003) calls them "traditional" valuation methods. To this group belong comparable method, investment/income method, profits method, development/residual method, contractor's method/cost method, multiple regression method and stepwise regression method. The second group, so called "advanced" valuation methods, try to analyse the market by direct mimicry of the thought processes of the players in the market and estimate the point of exchange (Pagourtzki, 2003). They are usually more quantitative and include the following methods: ANNs[7], hedonic pricing method, spatial analysis methods, fuzzy logic and ARIMA[8].

In the following subchapters, the most frequently used real estate valuation methods are described. The hedonic pricing method is handled in detail in the whole next chapter.

[5] A person qualified by education, training and experience to provide appraisals
[6] A public official responsible for valuing properties for tax purposes
[7] Artificial neutral networks
[8] Autoregressive integrated moving average

### 1.3.1 Comparable method

Comparable method is the simplest method, which can be used to directly determine the capital value. However, comparable method is rarely appropriate because of heterogeneity of the housing market. It can be useful only in case of sub-markets with a high degree of similarity, such as residential markets (Pagourtzki, 2003).

### 1.3.2 Investment/income capitalization method

Investment (income) method is similar to methods used for financial valuation, securities analysis or bond pricing. The value of **income producing** real estate is estimated, based upon the expectation of future benefits. It is related to both "market rent" that a property can be expected to earn and the "reversion"[9] when a property is sold (PropEx PLS, 2003). Anticipated cash flows are converted into present value by capitalizing[10] net operating income[11] (NOI) by so called capitalization rate (also known as cap rate), which is a rate of return on a real estate investment property derived by the market. In appraisal practice, it is extracted from sales of similar investment properties (PropEx PLS, 2003). Based on the present value, the investors find out how much they are willing to pay for the property.

The most common ways to estimate value by capitalization are called direct capitalization and yield capitalization. Which of these two methods is better, depends on several factors, the most important are timing and regularity of the cash flows, period of time the investment is held and whether or not long term leases are involved (PropEx PLS, 2003).

Direct capitalization is, because of its simplicity, a widely used approach. To estimate value of the property with this method, it suffices to divide its (annual) NOI by the market capitalization rate. One of the advantages of direct capitalization is that explicit projections of income are not required. However, it can be only used when future cash flows are not expected to vary significantly over time, so that the property's stabilized NOI can be estimated (PropEx PLS, 2003). This estimate is based on market data (e.g. rental rates, vacancy and collection loss rates or operating expense data) of comparable properties in the market area (Etter, 1994). Typically, either a single year's

[9] Any future interest kept by a person who transfers property to another
[10] A process of converting income to value
[11] Net operating income = all revenues from the property – all reasonably necessary operating expenses

income or average of several years' incomes is analysed. Another assumption is that the cap rate is a constant, i.e. it does not change over time, which could happen, for example, when the risk levels were expected to change (Etter, 1994).

In contrast to direct capitalization, yield capitalization requires not only explicit projections of income, but also holding period[12] as well as the property's reversion. In most cases, it considers the income streams for several years. The future benefits, i.e. any series of periodic incomes (with or without a reversion), are converted into present value by applying appropriate yield rates (discount rates) to various cash flows (PropEx PLS, 2003).

The most often used form of yield capitalization is so called discounted cash flow capitalization (DCF), which is analogous to net present value estimation in finance. This method requires estimates of each year's NOI along with the expected reversion value of the property at the end of the analysis period. The expected future cash benefits are then discounted to obtain the market value estimate (Etter, 1994). The advantage of DCF, compared to direct capitalization, is that it allows the buyer to reflect their expectation of changing NOI over time. This is useful, for instance, if the property has significant vacancy at the time of the appraisal, which is supposed to decrease in the future. Although the appraiser could use a market vacancy rate[13] instead of the property's actual vacancy rate in this case, it would result in a larger NOI of the property which could overstate its value. (Direct capitalization could also understate real NOI, for example when the future NOI would be expected to increase because of greater demand for space that would lead to higher rental rates. As already mentioned, direct capitalization namely does not take into account future projections of NOI.) To sum up, it is definitely better to use DCF rather than capitalization method in case that net operating income is expected to fluctuate over time, because it permits annual adjustments, so the estimate of NOI should be more realistic. On the other hand, if the expected changes in net operating income are not significant, both methods should produce the same result (Etter, 1994).

Now it begs the question, how to determine both the capitalization rate and the discount rate. In fact, there is a need for appropriate comparable sales in both cases. Estimating the market capitalization rate requires only NOI and the reported sales price

[12] The real or expected period of time during which an investment is attributable to a particular investor

[13] The percentage of all available units in a rental property that are vacant or unoccupied at a particular time

of each comparable property. If the market capitalization rate is derived from sales of properties whose vacancy rates as well as buyer expectations about becoming fully leased are similar to the subject, its value can be estimated with this market capitalization rate and unadjusted NOI (Etter, 1994). When appropriate comparables cannot be located, the appraiser must estimate the NOI and develop the capitalization rate from the best available comparables. Deriving a discount rate from the market is more complicated. One possibility is to determine the buyer's expectations about future NOI and reversion of each comparable property and, based on this information, calculate the discount rate for each property, which equates the expectations to the purchase price. After that, the appropriate discount rate for the subject property is selected, using the same process as in case of direct capitalization (Etter, 1994). However, this procedure would be difficult in practice. Therefore, the appraisers often estimate the discount rate by adding a risk premium and a liquidity premium to a relatively risk-free rate of return (e.g. U.S. Treasury bill rate). The problem is that ascertaining of the risk and return differences between the subject property and the security is quite complicated as well. Another approach could be to use investor's surveys reporting their expectations for several different property types; however, these expectations obviously tend to be general, which is not what we want.

The appraisers sometimes try several discount rates and the one that results in a market value estimate which is approximately the same as that obtained with direct capitalization is selected. Its reasonableness is then verified by comparing it to investor returns in other markets. Etter (1994) claims that "*in a perfect world, the discount rate is equal to the capitalization rate plus the weighted average of the net operating income and the property value annual growth rates*".

### 1.3.3 Profits method

Profits method is generally applied to **commercial** properties whose value is driven by the profitability of the businesses that occupy the buildings which are valued, such as hotels, guest houses, pubs, restaurants, leisure centres, cinemas or theatres (Pagourtzki, 2003). To be able to use this method, a necessary condition is that the subject property itself has an operational business currently running from within it (Investment Property Partnership, 2015). Many investors prioritize the profits method against the comparable method, because even if properties are in the same location and similar in size, appearance and quality of construction, there are still many factors which can cause the properties to have differing profit generating characteristics.

To calculate the value of a property using the profits method, the key financials derived from the occupying business must be established firstly. The appraisers need to obtain the financial accounts of the business for at least the last 3 years and to carefully examine them (Investment Property Partnership, 2015). The accounts should be accurate and reliable to allow them quick identification of the financial stature of the business (both currently and historically). Firstly, the gross profit is calculated as the difference between gross earnings (the total yearly revenue the business generates) and purchases (the materials that need to be bought so that the business can exist and perform its daily activities). After that, the appraisers calculate the net profit as the difference between gross profit and working expenses (expenses that occur daily and are integral to the running of a business, such as telephone, water, gas, electricity or business rates[14]). The net profit is a financial figure which shows how profitable the business really is and, along with investor's own judgement, research and investment appraisal should suffice to make an accurate investment decision (Investment Property Partnership, 2015).

### 1.3.4 Development/residual method

The residual method of property appraisal enables to estimate the value of plots or sides that can be developed. According to Pagourtzki (2003), the best manner of estimating site value is through comparable vacant land sales. The value of the site should be estimated as if it were vacant and available for its highest and best use. Each comparable sale should be described and necessarily include the data about location, grantor[15], grantee[16], recording data, date, sale price, financing, units of comparison, lot dimensions, configuration and size, physical and topographical characteristics, zoning, utilities and environmental influences. The estimate of the market value of the land in a redeveloped form, which is made either by comparison or by the investment method, is called gross development value. By deducting all costs incurred to put the property into the form that will command that price (demolition of the existing building, infrastructure works, construction costs, professional fees, finance cost and profit

[14] Tax on business properties
[15] The creator of a trust, meaning the individual whose assets are put into the trust
[16] An individual to whom a transfer or conveyance of property is made

required by the developer), a residue is produced, which represents the maximum capital expenditure for buying the land (Pagourtzki, 2003).

### 1.3.5 Contractor's/cost method

In some cases, the already mentioned valuation methods (comparable, investment, profits and residual) are not suitable for a particular property. The reason is that some buildings are extremely specialised in their nature, very rarely change hands on the open market and were not designed for commercial use as well. These buildings could have been designed, for example, for public sector, health care, military workers or Town Councils (Parker, 2011). Therefore, the contractor's method assesses the market value of the building by reference to its replacement cost[17]. The instructions are described by Pagourtzki (2003). Firstly, the appraiser must assess the market value of the raw land, by reference to comparable land values in an appropriate alternative use. After that, the cost of rebuilding a new building which could perform the function of the existing structure is added to this value. Finally, to allow for obsolescence and depreciation of the existing building relative to the new hypothetical unit, subjective adjustments are made. Pagourtzki (2003) claims that as it is reasonable to assume this mirrors the thought process of the owner-occupier, contractor's method should be viewed as a valid method of valuation. However, Parker (2011) argues that the assumption that the property values will be the same as the cost is flawed, because cost is an exact figure, whereas value being subjective is not.

[17] "*How much would it cost to replace the property, if the business were deprived of its use?*" (Pagourtzki, 2003)

# 2 Hedonic pricing model

Today, many real estate owners as well as investment fund managers still wonder how to properly value real estate assets. The market value of many assets is namely driven down by recent valuations based on comparable properties and area cap rates, although the underlying fundamentals have not changed. Moreover, some people claim that traditional discounted cash flow method also fails, because it uses broad-based assumptions such as discount rates, going-out cap rates, the effect of market cycles on disposition values[18] and so on (Monson, 2009). According to Monson (2009), the market value of real estate has far outweighed its true (intrinsic[19]) value. In his opinion, it is related to the fact that proper underwriting fundamentals seem to have been lost at the height of the latest real estate cycle as a consequence of irrational exuberance driven by cheap capital chasing deals. Nevertheless, Monson claims that for proper valuation, it is necessary to understand the intrinsic value of a real estate asset and the characteristics that contribute to its potential transaction price, which can be calculated only by fastidious underwriting.

In this chapter, an alternative valuation method called hedonic pricing model is introduced. It is particularly useful when traditional discounted cash flow models cannot be used because of the absence of a market, when no comparable buildings exist or for non-income generating buildings (Monson, 2009). The brief history of hedonic models and description of the most frequently occurring housing characteristics they use for the valuation is included as well.

## *2.1 Theoretical foundation of the hedonic pricing model*

To explain what is undertaken in the hedonic analysis of housing markets, Sheppard (1998) compares people using this method to private investigators or market researchers who study the demand for food without the possibility to enter the local grocer. However, they are allowed to photograph shopping baskets which provides them almost accurate information about the food each customer has purchased (not entirely accurate, because some items may be obscured in the shopping basket). They also see

[18] Disposition value assumes a shorter than average marketing time, due to the fact that the seller is under pressure to sell relatively quickly

the total cost of all items combined and are able to find out each shopper's income. From this information, they should be able, for example, to derive the demand for milk or infer how many customers would be willing to pay for removing of sugar import quotas.

In case of housing market, the situation is analogous. Each house namely consists of many unique characteristics and all of them may affect its value. The characteristics can be either physical (e.g. square footage, bathrooms and age) or not (e.g. school quality). They may be divided into two basic groups – internal, such as number of rooms, and external, such as distance from the nearest bus stop (Monson, 2009). Some of these characteristics can be valued differently across different geographical areas (for example, a swimming pool is likely to have greater value in a warmer climate). To make the valuation even more complicated, a certain house with a given set of characteristics may be valued differently by different buyers as a consequence of unique utility functions of individual people (Sirmans et al., 2005).

It is not possible to observe prices of individual characteristics directly. Nevertheless, (somewhat imprecise) observations of what attributes are purchased when buying a particular dwelling, reasonable good observations of what is spent for it and the income of the household are available. Using this information and the hedonic regression analysis, which allows the total housing expenditure to be broken down into the values of individual housing characteristics (and thereby enables to calculate the monetary value[20] of each characteristics by observing the differences in the market price of housings that share the same attributes), we should be able to estimate their marginal contribution and answer some important questions (Sheppard, 1998).

Using the hedonic pricing model assumes that the customer's utility is derived by the housing characteristics, the value of this utility can be priced (and the sum of the prices of all characteristics of a particular dwelling is equal to the price the customers have paid for it) and in housing consumption, the customers pursue maximization of utility within their budget constraint (Malpezzi et al., 1980). The model can be generally expressed by the following equation:

$$P = f(x_1, x_2, \dots, x_n)$$

[19] The actual value of a company o ran asset based on an underlying perception of its true value including all aspects of the business, in terms of both tangible and intangible factors. This value may or may not be the same as the current market value

[20] The amount of value an item or service has in relation to if it were sold for cash to a willing buyer

This equation tells us that the (market) price of the house $P$ is a function of its characteristics $x_1, x_2, \ldots, x_n$. The partial derivatives of the price with respect to these characteristics $\partial P/\partial x_i$ show the marginal willingness to pay for an additional unit of characteristic (it can be both positive and negative), which enables to deduce the implicit price of each one of them (Morancho, 2002). Furthermore, using this information, it is possible to predict transaction prices of the dwellings based on the attributes they consist of and to determine which of these attributes most highly affect it, which is useful information especially for investors and developers (Monson, 2009). The observed selling price, which is generally believed to better minimize bias than other possible measures (such as an owner's self-assessment) is usually used as a proxy for the value of the house (Sirmans et al., 2005). One of the greatest advantages of the regression analysis is that it allows to measure characteristics that are qualitative in nature in quantitative terms (such as the already mentioned distance from the building to the nearest bus stop) (Monson, 2009).

Hedonic pricing method, as well as any other valuation methods, has some disadvantages. Firstly, the results are location-specific and therefore difficult to generalize across different geographic locations, which is the reason why the models are most often used to gain insight into workings of a particular market (Sirmans et al., 2005). Secondly, the studies using hedonic models may define and measure the characteristics differently (for example, bathrooms may be measured either simply as the number of bathrooms or using binary variable) which complicates comparison of the individual studies. The functional forms can differ as well. The model is often estimated using semi-logarithmic form (instead of linear form) with the natural log of price which enables to calculate the percentage change in price for a one-unit change in a given characteristics and thereby allows for variation in characteristic prices across different price ranges within the sample. Moreover, it helps to minimize the problem of heteroscedasticity (Follain and Malpezzi, 1980). Another disadvantage could be that the number of characteristics that can be included in the model is almost limitless and some of them are highly correlated with each other, which can cause estimation problems. For instance, if a correlation coefficient of fitness centre is negative, it does not necessarily mean that including this characteristic into a building will lower its transaction price. Another possible explanation may namely be that at the time when the dwellings from the data set were transacted, fitness centres were not highly sought after by building owners and therefore viewed as a lost leasable space or that the owners of some ramshackle buildings had built the fitness centres there as an attempt to increase their attractiveness and value, for example (Monson, 2009). However, even if these drawbacks are taken into account, hedonic modelling can still be useful in addressing many issues in housing valuation. It has already been used many times not only in valuing “obvious” characteristics such as square footage but has also been useful in measuring the effect of other more interesting components, including school quality, proximity to high voltage lines and percentage of Blacks or Hispanics in an area (Sirmans et al., 2005).

## 2.2 Early history of hedonic models

It is not easy to identify the “father” of hedonic modelling. According to Malpezzi (2002), a study called *Hedonic Price Indexes with Automobile Examples* written by A. T. Court in 1939 is often cited as the beginning of hedonic models, although it had nothing in common with a housing application (in fact, the study is about hedonic price index for automobiles). Court was an economist for the Automobile Manufacturers’ Association between 1930 and 1940 who found out that a single variable could not explain demand for automobiles. He used three independent variables to explain the price of automobile: dry weight, wheelbase and horsepower (Sirmans et al., 2005). Moreover, Court used a semi-log form of the model, which is even today considered modern.

However, although Court was the first who used the term “hedonic”, Colwell and Dillmore (1999) highlight that hedonic model was applied already in 1922 by G. C. Haas to estimate the value of a farmland. Furthermore, the similar study was made again four years later, in 1926, by H. A. Wallace. Colwell and Dillmore (1999) claim that Court has just developed their ideas for a hedonic model from discussions with the chief of the Bureau of Labour Statistics, who probably knew about the work by Wallace (and maybe the work by Haas).

Later studies important for hedonic modelling were written in 1966 by K. J. Lancaster, in 1971 by Z. Griliches and in 1974 by S. Rosen. Lancaster developed a sophisticated branch of microeconomic theory in which utility is generated not by the goods itself, but by characteristics of the goods and thereby he provided microeconomic foundation for estimating the value of utility-generating characteristics, which is applicable not only to housing, but also to financial assets, the labour-leisure trade-off and the demand for money (Malpezzi, 2002). Rosen dealt with characteristics as well, but he concentrated rather on the way how suppliers and consumers interact within a framework of bids and offers for characteristics (Malpezzi, 2002). Furthermore, although he did not discuss functional form explicitly much, his work provided the basic foundation of nonlinear hedonic pricing models (Sirmans et al., 2005). Griliches, as well as Court, applied hedonic model in automotive. According to Morancho (2002), the theoretical support for the development of the hedonic models was provided by Griliches and Rosen.

## 2.3 Typical characteristics by category

Sirmans et al. (2005) have examined approximately 125 studies estimating hedonic pricing models from a number of different journals (including *Journal of Real Estate Research*, *Journal of Real Estate Finance and Economics*, *Real Estate Economics*, *Journal of Urban Economics*, *Land Economics* or *The Appraisal Journal*) published between 1995-2005, which enabled them to identify most often used housing characteristics by eight categories:

- **Construction & Structure**: Lot size, square feet, age, number of bathrooms and number of bedrooms;
- **House Internal Features**: Full baths, half baths, fire place, air conditioning, hardwood floors and basement;
- **House External Amenities**: Garage/garage spaces, deck, pool, porch and carport[21];
- **Environmental – Natural**: Lake front or view, ocean view and "good view";
- **Environmental – Neighbourhood & Location**: Location (generally measured as a neighbourhood identifier or zip code), crime (usually measured as the crime rate for a given area), distance (typically measured as distance from the city centre), golf course (usually measured as being on or near a golf course) and trees (usually mean a wooded lot versus an open lot);
- **Environmental – Public Service**: School district, percentage minority in school district and access to a public sewer;
- **Marketing, Occupancy & Selling Factors**: Assessor's judgement of quality, the assessed condition of the house, whether the house is vacant at the time of sale, whether the house is owner-occupied, time on market and time trend;

[21] A covered structure used to offer limited protection to vehicles, primarily cars, from the elements

- **Financial issues**: Type of financing, whether the house is in foreclosure and property taxes.

# 3 Empirical analysis

In this chapter, the hedonic pricing model, which was described in detail in the previous chapter, will be used to model the prices of Prague flats. Firstly, the theoretical framework I am going to use will be discussed. After that, the models and estimation methods as well as the data used will be described. Finally, after verifying all necessary assumptions, the results will be presented.

## *3.1 Theoretical framework*

As already discussed in the previous chapter, one possibility how to determine the value of real estate is to sum up the monetary values of its characteristics, which can be estimated by the hedonic pricing model, typically using ordinary least squares (OLS) estimation method. However, it is not clear from the theory whether these values are constant or take form of percentage change in the price. Therefore, three basic forms of the model (level-level, log-level and log-log), which are most often used, will be estimated. The results will be compared with literature using similar models in other locations.

The log-level model, which is used probably most frequently, will be used as a base for other investigation. In attempt to explain the variance in the prices of flats in Prague better, a hypothesis about possible quadratic relationship between the price and distance from the dwelling to the city centre will be tested. The assumption is that that the price should decrease with increasing distance from the city centre (e.g. because of nice view on the historical buildings or proximity to shops and services), but only until some point and then it may start to increase again, as some people prefer to live close to nature rather than close to the centre.

Similar hypothesis will be tested about the relationship between the price of the dwelling and its usable living area (size). It is clear that the price will most likely increase with increasing size. However, I do not think that the price increases exponentially, as the log-level form of the model suggests, because the price per square meter tends to be higher in case of smaller flats. Therefore, I will include also the squared size in the model and check whether it is significant and its sign (I expect it will be negative, which would mean that with increasing size, the percentage increase in price decreases).

Another issue that will be investigated is the influence of the floor on which the flat is located on its price. Intuition offers two possible effects of the floor. The first effect is positive as in case of upper floors, there may be a better view and/or less noise from the street. The second effect is negative, as it is inconvenient to climb to the upper floors. Obviously, the second effect is eliminated by the presence of an elevator.

It is reasonable to expect that some pattern of heteroskedasticity will be present in all models. The expectation is that the higher the prices, the higher the variance is. Therefore, I will test for heteroskedasticity (using Breusch-Pagan test) and if it is present, I will correct for it using heteroskedasticity-robust standard errors. In the final hedonic pricing model, which will be based on the results of the three above described hypotheses, the form of heteroskedasticity will be specified and weighted least squares (WLS) method, which is (provided that we have correctly specified the form of the variance, as a function of explanatory variables), more efficient than OLS, although no longer unbiased when weights have to be estimated using the same data (Wooldridge, 2003), will be used. The results of OLS and WLS will be interpreted and compared and their reasonableness will be discussed.

## *3.2 Econometric model*

For the purpose of comparison of the results with related literature in other locations, OLS was chosen as it is the most commonly used estimation method. As mentioned in the previous subchapter, three different forms of this method will be used. Their interpretation is described below (price is a dependent variable, $x_i, i = 1, \ldots, k$, are explanatory variables and $\varepsilon$ is an error term; $\ln$ indicates natural logarithm):

- **Level-level**: $price = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$

  Interpretation: Holding all other factors affecting price fixed, if $x_i, i = 1, \ldots, k$, changes by one unit, the price will change by $\beta_i$ units.

- **Log-level**: $\ln(price) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$

  Interpretation: Holding all other factors affecting price fixed, if $x_i, i = 1, \ldots, k$, changes by one unit, the price will change by $100\beta_i$ % .

- **Log-log**: $\ln(price) = \beta_0 + \beta_1 \ln(x_1) + \cdots + \beta_k \ln(x_k) + \varepsilon$

  Interpretation: Holding all other factors affecting price fixed, if $x_i, i = 1, \ldots, k$, changes by 1 %, the price will change by $\beta_i$ %.

When testing potential quadratic relationship between price and distance of the dwelling from the centre (cendis), the following model will be used:

$$\ln(price) = \beta_0 + \beta_1 cendis + \beta_2 cendis^2 + \beta_3 x_3 + \cdots + \beta_k x_k + \varepsilon$$

Interpretation: $\frac{\partial \ln(price)}{\partial cendis} = \beta_1 + 2\beta_2 cendis$, i.e. the percentage change in price is not constant as it depends on the actual level of cendis. The model for testing the (exponential) relationship between price and size will be analogous.

When testing the relationship between price and floor of the building on which the flat is located, the effect of variable floor on price is assumed to be a combination of "climbing" effect and "view_and_noise" effect, where climbing and view_and_noise are unobserved effects depending on the floor. Therefore, we can write the model in this way:

$$\ln(price) = \beta_0 + \gamma_1 climbing + \gamma_2 view\_and\_noise + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

In case there is no elevator in the building, floor is used as a proxy variable for both climbing and view_and_noise:

$$\ln(price) = \beta_0 + (\gamma_1 + \gamma_2) floor + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \ if \ elevator = 0$$

Provided that there is at least one elevator in the building, climbing is no longer necessary and $\gamma_1$ is assumed to be 0:

$$\ln(price) = \beta_0 + \gamma_2 floor + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \ if \ elevator = 1$$

To estimate these effects, a semi-log model with an interaction term is used:

$$\ln(price) = \beta_0 + \beta_1 floor + \beta_2 elevator + \beta_3 elevator * floor + \beta_4 x_4 + \cdots + \beta_k x_k + u$$

The partial effect of elevator can be also calculated as $\frac{\partial \ln(price)}{\partial elevator} = \beta_2 + \beta_3 floor$. The term $\beta_3 floor$ allows the partial effect of elevator to change with the floor.

The partial effect of floor is calculated as $\frac{\partial \ln(price)}{\partial floor} = \beta_1 + \beta_3 elevator$. The term $\beta_5 elevator$ allows the partial effect of floor to change (presumably increase) when the elevator is present.

Therefore, if elevator = 0,

$$\frac{\partial \ln(price)}{\partial floor} = \beta_1 = \gamma_1 + \gamma_2$$

And if elevator = 1,

$$\frac{\partial \ln(price)}{\partial floor} = \beta_1 + \beta_3 = \gamma_2$$

The estimate of $\gamma_2$ is equal to the sum of estimates $\beta_1$ and $\beta_3$ and the estimate of $\gamma_1$ is equal to $\beta_1 - (\beta_1 + \beta_3) = -\beta_3$.

When dealing with heteroskedasticity, Breusch-Pagan test is run automatically (i.e. not manually) by STATA[22]. The same holds for heteroskedasticity-robust standard errors. WLS regression is performed manually, based on the procedure described in Wooldridge (2003): Firstly, a regression of $\ln(price)$on $x_1, x_2, \ldots, x_k$ is run and the residuals $\hat{u}$ are obtained. Secondly, $\ln(\hat{u}^2)$ is created by squaring the OLS residuals ant taking the natural log of them. After that, the regression of $\ln(\hat{u}^2)$ on $x_1, x_2, \ldots, x_k$ is run, the fitted values $\hat{g}$ are obtained and subsequently exponentiated: $\hat{h} = \exp(\hat{g})$. Finally, the original equation is estimated by WLS, using weights $1/\hat{h}$.

The choice of the explanatory variables was not easy. Based on similar studies in other countries, I tried to include as many variables that typically appear in hedonic pricing models as possible and also some less usual variables, such as distance from the dwelling to the nearest ATM, which could be a proxy for shops and services as ATMs are typically located on main streets, or a dummy variable indicating whether there is a kitchen or kitchenette in the flat, to find out what is more popular. I also wanted to include the distance from the dwelling to the nearest underground station, but a high multicollinerity between this variable and distance to the city centre was detected, so I had to eliminate it. The list of the variables that were finally used is described in the next subchapter.

## 3.3 Data description

In our analysis, the following housing characteristics were used:

- **ATM** - the distance from the dwelling to the nearest cash dispenser (in meters);
- **CENDIS** - the linear distance from the dwelling to the centre of Prague, measured as the minimum of the linear distances from the dwelling to the Saint Wenceslas Statue, entrance to the Old Town City Hall and the Powder Tower (in meters); the distances were determined using VBA[23] and Google Maps[24] and may be slightly imprecise as the street numbers were not available;

[22] The statistical software package STATA 12.0 is used
[23] Visual Basic for Applications

- **ELEVATOR** - a dummy variable that takes a value of 1 when the building in which the flat is located has an elevator and 0 otherwise;
- **FLOOR*** - the floor of the building in which the flat is located;
- **FOOD** - the distance from the dwelling to the nearest restaurant/grocery;
- **KITCHEN** - a dummy variable that takes a value of 1 when the dwelling has a kitchen and 0 when it has only a kitchenette[25];
- **mat_BRICK -** a dummy variable that takes a value of 1 when the building in which the flat is located is brick and 0 otherwise (*this variable, together with the following three variables starting with "mat_", describes the material of the building in which the flat is located, variable mat_BRICK was set as a base*);
- **mat_MIXED -** a dummy variable that takes a value of 1 when the building in which the flat is located is mixed and 0 otherwise;
- **mat_PANEL -** a dummy variable that takes a value of 1 when the building in which the flat is located is panel and 0 otherwise;
- **mat_SKELETON -** a dummy variable that takes a value of 1 when the building in which the flat is located is skeleton and 0 otherwise;
- **own_COOPRTV -** a dummy variable that takes a value of 1 when the dwelling is a cooperative property and 0 otherwise (*this variable, together with the following two variables starting with "own_", describes the type of ownership of the dwelling, variable own_PERSONAL was set as a base*);
- **own_PERSONAL -** a dummy variable that takes a value of 1 when the dwelling is a personal property and 0 otherwise;
- **own_STATE -** a dummy variable that takes a value of 1 when the dwelling is a state-owned/municipal property and 0 otherwise;
- **PHARMACY** - the distance from the dwelling to the nearest pharmacy (in meters);
- **POST** - the distance from the dwelling to the nearest post office (in meters);
- **PRICE*** - the offer price of the dwelling (in thousands of Czech crowns);[26]
- **ROOMS** - number of rooms in the dwelling;

---

[24] URL: https://www.google.cz/maps?source=tldsi&hl=cs

[25] A small kitchen or part of another room equipped for use as a kitchen

[26] It is important to bear in mind that since only the information about the offer prices was available, the realized prices may differ as they may be subject to bargaining, which can cause the results of our analysis to be slightly inaccurate

- **SIZE** - (usable) floor area of the dwelling (in square meters);
- **stt_AFTERREC** - a dummy variable that takes a value of 1 when the building in which the flat is located is after reconstruction and 0 otherwise (*this variable, together with the following six variables starting with "stt_", describes the status of the building in which the flat is located and, variable stt_GOOD was set as a base*);
- **stt_BEFORREC** - a dummy variable that takes a value of 1 when the building in which the flat is located is before reconstruction and 0 otherwise;
- **stt_GOOD** - a dummy variable that takes the value of 1 when the state of the building in which the flat is located is "good" and 0 otherwise;
- **stt_NEW** - a dummy variable that takes a value of 1 when the building in which the flat is located is new (or it is a project) and 0 otherwise;
- **stt_POOR** - a dummy variable that takes a value of 1 when the state of the building in which the flat is located is "poor" and 0 otherwise;
- **stt_UNDERCON** - a dummy variable that takes a value of 1 when the building in which the flat is located is under construction and 0 otherwise;
- **stt_VERYGOOD** - a dummy variable that takes a value of 1 when the state of the building in which the flat is located is "very good" and 0 otherwise.

*Notes:

- In case of variable **FLOOR**, there were eighteen observations with an underground floor (of which sixteen flats were in the first underground floor and two flats in the second underground floor). One possibility how to deal with these data could be to replace the negative values with positive values, (people do not care whether they go upstairs or downstairs when getting to their flat, because they obviously have to go out as well). However, higher floor connotes also better view (or less noise), and the view from the second floor is obviously not the same as from the second underground floor. The interpretation of negative floors is also unclear and beyond the scope of this text, so I decided to drop these sixteen observations from the sample;
- In case of variable **PRICE**, two outliers were identified (see Appendix 2): 75.000 and 103.200 (in CZK thousands). Both observations were dropped from the sample. The reason is that based on the data set, I believe the value of

103.200 CZK thousands is an error, because none of the explanatory variables (including variables that are not the part of our analysis) apparently explain why it should be so high; furthermore, the usable floor area of this flat is only 80 square meters, so even if the price was 10.320 CZK thousands, it would still seem too high to me, compared with other buildings with similar characteristics. In case of the value of 75.000 CZK thousands, I believe it can be true, because, according to the description, the flat which costs this amount seems very attractive. However, the variables that probably mostly contribute to the fact that the price is so high, such as view on the Prague Castle or its fully furnishing with elements of modern art and Asian culture as well as antiquarian furniture, are not included in our analysis, so I decided to drop this variable as well.

The sample has 5996 observations in total. The simple summary statistics (minimum, maximum, mean and variance) as well as the number of observations of each individual variable are reported in the table in Appendix 1. It is also clear from this table that some data are missing. STATA keeps track of them and simple ignores the appropriate observations when computing a regression. According to Wooldridge (2003), if the data are missing at random, which is supposed in this text, there is no statistical consequence of it (apart from the reduced random sample from the population).

## *3.4 Empirical results*

### 3.4.1 Assumptions and basic provisions

Firstly, it is important to show the crucial assumptions are met by the data, so that the analysis can be trusted. Our data are cross-sectional, which implies that they should satisfy the following six assumptions (Wooldridge, 2003):

- MLR[27].1 (**Linear in parameters**) - the models are linear in parameters (see Econometric model section);
- MLR.2 (**Random Sampling**) - the data were downloaded from a reality server Sreality.cz. Since both individuals and real estate brokers can place an ad there and Sreality.cz is the most visited reality server in the Czech Republic (at least it

[27] Multiple linear regression

is written on their website), I believe that everybody who wants to sell a flat in Prague has a great motivation to place an add there, so the sample of observations is supposed to be random;

- MLR.3 (**No Perfect Collinearity**) - the statistical software package STATA, which is used for the analysis, would not allow to violate this assumption; in case of dummy variables, one of them was always stated as a base (see the Data description section);
- MLR.4 (**Zero Conditional Mean**) - unfortunately, it is not possible to test whether this assumption holds (Wooldridge, 2003); however, as many variables as possible were used to maximize the probability that the explanatory variables will not be correlated with the error term;
- MLR.5 (**Homoskedasticity**) - it was tested for heteroskedasticity and found out that the heteroskedasticity is present, so it was subsequently either corrected for it using heteroskedasticity-robust standard errors or an alternative model (WLS) was used;
- MLR.6 (**Normality**) - the sample is large enough, so, according to Wooldridge (2003), this assumption can be dropped.

MLR.1 – MLR.4 are used to establish unbiasedness of OLS, MLR.5 was added to derive the usual variance formulas and to conclude that OLS is the best linear unbiased estimator and MLR.6 was added to obtain the exact sampling distributions of t statistics and F statistics, so that exact hypotheses tests can be carried out (Wooldridge, 2003).

When estimating the models, we are primarily interested in the p-value of each variable, which is the lowest significance level on which the null hypothesis that the variable's coefficient is equal to 0 can be rejected. If the p-value is greater than the desired significance level, we say that the variable is not significant. In our case, the desired significance level was set to 5 %. Insignificant variables are always tested for joint significance using the F test and if the null hypothesis that all their coefficients are equal to 0 cannot be rejected, the variables are dropped from the model. However, in case of dummy variables regarding the material or status of the building in which the flat is located or the type of ownership of the dwelling, if at least one variable of the group is significant, the other variables from this group are not dropped from the model as the interpretation of the significant variable would not be clear. The p-values of these insignificant variables have red colour in the tables with results.

### 3.4.2 Basic model(s) and the comparison

All basic models (level-level, log-level as well as log-log) were estimated with entire set of explanatory variables and tested for heteroskedasticity, which was present in all three cases. To correct for it, heteroskedasticity-robust standard errors were calculated. After that, non-significant variables were tested for joint significance and subsequently dropped (FOOD and PHARMACY in case of level-level model, FOOD, PHARMACY and KITCHEN in case of log-level model and lATM and lPHARMACY in case of log-log model, where l indicates natural logarithm). Finally, new models (without the non-significant variables) were estimated. The results, including both robust and "usual" standard errors for comparison, are presented in Appendix 3 (level-level model), Appendix 4 (log-level model) and Appendix 5 (log-log model). The signs of significant variables are mostly as expected, except for the variable FOOD which was not dropped only from log-log model and here it has positive sign, indicating that if the distance from the dwelling to the nearest restaurant/grocery increases by 1 %, the price of his dwelling will increase by 0.014 % (rounded to three decimal places), which is not so much. Possible explanation could be that FOOD is correlated with some other variables that are not included in the model. Much more surprising could be the fact that in all three models, the robust standard error of variable stt_POOR is much lower than usual standard error, which makes this variable significant with positive sign. That would indicate that the flat in a building whose state is good is more expensive than if the state was poor, which is obviously nonsense. However, in fact, only two buildings from the sample were reported to be in a poor state, so this result should not be taken seriously. The only reason why this variable was included in the model was an effort not to complicate the interpretation of other variables describing the status of the buildings.

As already mentioned, the main purpose of these three relatively simple models is to enable the comparison of their results with similar models in other three locations. That is also the reason why the results will not be described in detail here.

The first location is the city of Castellón in Spain. Morancho (2002) gathered 810 observations and ran a hedonic regression, whose main purpose was to estimate the value of urban green areas. However, in his level-level model appeared four variables that were used in our models as well – CENDIS, ELEVATOR, FLOOR and SIZE. FLOOR was unfortunately not significant at the 5 % significance level. The coefficients of other variables were (-0.08649), 444.055036 and 17.95148, respectively.[28] Their signs are the same as in our model, but the magnitudes differ – the coefficient of CENDIS is approximately three times higher in our model and the same holds for SIZE, whereas the coefficient of ELEVATOR is somewhat higher in Castellón. Morancho used the variable SIZE as the only explanatory variable in a log-log form of the model as well with the resultant coefficient 1.06. To make it 100 % comparable, I ran the same simple regression and obtained a coefficient 1.03, so the difference is minimal in this case.

The second location is Marion County (Indianapolis) in Indiana. Ottensmann et al. (2008) used the data on 8.722 house sales recorded in the MLS[29] database for the year 1999. Three variables that were used in their log-level model have been used in our model as well: CENDIS, ROOMS and SIZE. Variable CENDIS was non-significant this time. The coefficients of variables ROOMS and SIZE were 0.0131 and 0.00226044, respectively. Both estimates have the same signs as in our model, but their magnitudes differ quite a lot – the coefficient of ROOMS is almost twelve times higher in our model and the coefficient of SIZE is approximately three times higher.

The third and last location is Israel. Eshet et al. (2007) primarily wanted to investigate the economic value of externalities related to waste transfer stations in Israel. The variables that appeared in our as well as their models are ELEVATOR and SIZE. In case of log-log model, the estimated coefficients were 0.132 and 0.647, respectively. The coefficient of ELEVATOR is more than two and half times higher than in our model, whereas the coefficient of SIZE is just a little bit lower than that in our model. In case of level-level model, the coefficients are equal to 29.86760255 and 606.7588836, respectively.[30] The coefficient of SIZE is almost two times higher in our model, whereas the coefficient of ELEVATOR is almost two times lower in our model.

### 3.4.3 Testing hypotheses

When investigating the relationship between lPRICE and CENDIS, the model was estimated with entire set of explanatory variables plus a new variable CENDIS2,

[28] The prices were expressed in pesetas, so the last known exchange rate from the year 2001 (19.22 CZK = 100 ESP) was used to make the estimates comparable

[29] Multiple Listing Service

[30] [30] The prices were expressed in US dollars, so the exchange rate (1$ = 29.7339 CZK) was used to make the estimates comparable

which is a squared CENDIS. Subsequently it was tested for heteroskedasticity, which was present, so the heteroskedasticity-robust standard errors were calculated. The non-significant variables were not dropped from the model this time, since we are now interested only in the variables CENDIS and CENDIS2. The table with the results (see Appendix 6) shows that both variables are significant, which confirms our hypothesis that there is a quadratic relationship between lPRICE and CENDIS. It is supported also by the graph in Appendix 7, except for the fact that in subchapter 3.1, we expected the price to increase with the distance after some point, which is not this case. However, the signs of both estimates are as expected.

The investigation of the relationship between lPRICE and SIZE was analogous. Again, the model was estimated with entire set of explanatory variables plus a new variable SIZE2, which is a squared SIZE. The heteroskedasticity was present again, so the heteroskedasticity-robust standard errors were used to calculate the p-values. Again, both SIZE and SIZE2 showed to be significant (see Appendix 8), which confirms the hypothesis that the relationship between lPRICE and SIZE is not exponential. The graph in Appendix 9 supports it. The signs of both estimates are as expected again.

In case of investigation of the effect of ELEVATOR on lPRICE, a new variable called ELEVFLOOR, which is a product of variables ELEVATOR and FLOOR, was added into the model. The procedure was similar to the previous two models and the results are presented in Appendix 10. This time, we are interested only in the variables FLOOR and ELEVFLOOR. Unfortunately, both these variables are non-significant this time, which does not support our hypothesis that there is a synergy. The reason could be that as there are much more buildings with an elevator than without an elevator in the sample, the variables ELEVATOR and ELEVFLOOR are highly correlated.

### 3.4.4 Final model(s) and the interpretation

Based on the results of testing of the three hypotheses investigated in the previous subchapter, I decided to include variables CENDIS2 and SIZE2 into the log-level model in effort to better explain the variance in price. The final model was also estimated with entire set of explanatory variables, including CENDIS2 and SIZE2, and tested for heteroskedasticity, which was present as in every model. This time, not only the heteroskedasticity-robust standard errors were calculated, but also the form of heteroskedasticity was estimated, so that the model could be estimated using WLS as well. The results of OLS and WLS are presented in Appendix 11. Variable PHARMACY was dropped as it was highly non-significant.

The R-squared from OLS model is 0.8327, which is really higher than in the original log-level model, where it was “only” 0.7998, but this is probably caused to a certain extent by the fact that in the original model, not only PHARMACY, but also FOOD and KITCHEN were dropped. The R-squared from WLS model is 0.8756, but this is a natural consequence of the fact that the used weights are inversely proportional to the variance at each level of the explanatory variables. The signs of the estimates are the same regardless of which model is used, except for variable stt_BEFORREC, which is in case of OLS positive (but not significant) and in case of WLS negative (and significant at the 10 % significance level). Other variables that are not significant at the 5 % significance level are mat_SKELETON and own_STATE. Variable mat_MIXED is non-significant only in case of OLS.

Although WLS is not unbiased, as we had to use FGLS because the exact form of heteroskedasticity was not known, it is still consistent and asymptotically more efficient than OLS (Wooldridge, 2003). That is the reason why I believe more the WLS estimates and when interpreting the results, the OLS coefficients will be given in brackets only for comparison. The interpretation of the individual variables follows:

- **ATM**: If the distance from the dwelling to the nearest ATM increases by 1 kilometre, the price will decrease by 6.32 % (5.82 %). This change of price is most probably not caused only by the presence of ATM. I rather think the reason is that ATMs are usually located on the main streets with many shops and services, so ATM is likely to be something like a proxy for them. If this interpretation is correct, I believe that the coefficient is quite realistic;
- **CENDIS**: If the distance from the dwelling to the city centre increases by 1 kilometre, the price will change by (-0.0892+2*0.00000372*SIZE)*100 % = (-0.0892+0.00000744*SIZE)*100 %, i.e. it would decrease by 8.9 % for an average CENDIS, which is 4.9 km. The point where the price would start do increase with increasing distance from the city centre is at the distance of 11.9 km from the city centre. If OLS estimates are used for the calculation and the distance from the city centre increases by 1 kilometre, the price will decrease by 10.8 % for an average CENDIS and the point where the price would start to increase with increasing CENDIS is at the distance of 10.6 km from the city centre;

- **ELEVATOR**: If the building in which the flat is located has an elevator, its price will be 3.3 % (5.7 %) higher compared to the same flat which is located in the building without an elevator;
- **FLOOR**: If the floor where the flat is located increases by 1, the price will increase by 0.4 % (0.5 %);
- **FOOD**: If the distance from the dwelling to the nearest restaurant/grocery increases by one kilometre, the price will increase by 8.8 % (9.4 %). This result is somewhat surprising, as I expected the coefficient to be negative (I believe that it is an advantage for people to live close to some restaurant/grocery, since when they are hungry, they do not have to go far away). However, as the network of shops/restaurants in Prague is very dense, this variable probably will not affect the price so much;
- **KITCHEN**: If the flat has a kitchen, its price will be by 5.4 % (3.5 %) lower than the price of the some flat with kitchenette instead of kitchen. This suggests that in people's opinion, the kitchen takes us too much space which could be utilized more usefully. Therefore, modern flats are mostly built with a kitchenette and it is also seen in the sample that kitchenettes predominate;
- **MATERIAL**: If the building in which the flat is located is mixed, the price will be higher by 4.7 % (the OLS estimate is not significant) compared to the same flat which is located in a brick building. On the other hand, if it is panel, the price will be lower by 15.5 % (15.3 %);
- **OWNERSHIP**: If the flat is a cooperative property, the price will be lower by 13.2 % (15 %) compared to the same flat which is a personal property;
- **POST**: If the distance from the dwelling to the nearest post office increases by 1 kilometre, its price will decrease by 7.2 % (8.4 %). I believe that the interpretation will be similar as in case of ATM, because only the presence of post office probably would not influence the price so much;
- **ROOMS**: If the number of rooms changes by one, the price will increase by approximately 8.3 % in both models;
- **SIZE**: If the usable floor area of the flat increases by 1 metre square, its price will change by (0.0159675-2*0.0000345*SIZE)*100 % = (0.0159675-0.000069*SIZE)*100 %, i.e. it would increase by 1.05 % for an average SIZE, which is 79.9 metres square. The size from which the price would start do

decrease with increasing price of the flat is 231.4 metres square. If OLS estimates are used for the calculation and the SIZE of flat increases by 1 metre square, its price will increase by approximately 1.05 % for an average SIZE as well and the size from which the price would start to decrease with increasing price is 250.8 metre square;

- **STATUS**: As it is not entirely clear what it means when the building is in a "good" state or what is the difference between "good" and "very good" state, it does not make much sense to enumerate the percentage differences between the "good" building and other statuses. However, it may be useful to compare the coefficients (apart from variables stt_BEFORREC, which is non-significant in both models, and stt_POOR, whose interpretation, as already explained, should not be taken seriously). We can see that all other four coefficients have positive sign, which means that we would have to pay more for the flat located in the building whose status is different than good. If they are ordered based on the size of their coefficients ascendingly, it will look like this: stt_VERYGOOD, stt_AFTERREC, stt_NEW and stt_UNDERCON. I believe that this order is quite logical. Maybe it is a little bit strange that people are willing to pay more for flat located in the building that is under reconstruction than for new flat, but this can be associated for example with the fact that older flats can be (partly) equipped, which saves some money.

Finally, I would like to determine which of the "non-dummy" variables has the largest impact on price. One way how to make their coefficients comparable is to multiply the percentage change they cause in the price when they change by one unit by standard deviation of the appropriate variable's values. I came to an unambiguous conclusion that the largest percentage change in PRICE would be caused by variable SIZE, as the change in size by one standard deviation from mean value will cause the increase in price by 40.3 % (40.7 %).

# Conclusion

The purpose of this work was to model the prices of Prague flats, based on the public available data regarding their characteristics. The first part of the thesis dealt with the housing market. Its main participants as well as its special features that distinguish the housing market from standard commodity markets were described. Most attention was devoted to the discussion of the most frequently used valuation models of real estate – investment/income capitalization method, profits method, development/residual method and contractor’s/cost method.

The second part was focused on hedonic pricing model, which is another valuation method which can be used even if the other methods fail for some reason. This method allows the total housing expenditure to be broken down into the values of individual housing characteristics (e.g. square footage or number of rooms) and enables to calculate the monetary value of each of them by observing the differences in the market prices of housings that share the same attributes. The value of the house is subsequently calculated as the sum of the values of all its characteristics. The brief history of hedonic pricing method and the list of most frequently used housing characteristics were included in this chapter as well.

The last, most important part of this work was an application of the hedonic pricing model in practice. Firstly, simply level-level, log-level and log-log regressions were performed, using heteroskedasticity-robust standard errors to correct for heteroskedasticity. The effects of several variables were compared with the same variables that appeared in similar regressions performed in Spain, Indiana and Israel. Although the signs were the same, the magnitudes differed quite considerably, which could be partly caused by the fact that the other explanatory variables were different in both models. However, I believe the main reason is that the results are location-specific.

For the purpose of another investigation, the log-level form of the model was chosen. In effort to make the model more realistic, several hypotheses regarding the relationship between the price and selected explanatory variables were tested. It turned out that the relationship between price and distance from the dwelling to the city centre is quadratic and the relationship between price and the square footage of the dwelling is not exponential. Therefore, the squares of these two variables were included into the model. Another hypothesis related to the two possible effects of the floor on which the flat is located on its price. For our purposes, they were called “climbing” effect and “view_and_noise” effect. In order to be able to calculate them, the product of variables floor and elevator was added into the model. However, this interaction term was not significant, so our hypothesis that there is a synergy effect was not confirmed.

Based on our findings, a new log-level model was created and estimated by both OLS with heteroskedasticity-robust standard errors and WLS, after estimating the form of heteroskedasticity. The OLS and WLS coefficients of all explanatory variables were interpreted and compared. The signs of the coefficients of all significant explanatory variables were the same, regardless of the method which was used, and the differences in magnitudes were not large. It was found out, for example, that people prefer kitchenette to kitchen and are willing to pay 5.4 % (using the WLS estimate) more for the flat with kitchenette compared to the same flat with kitchen.

Finally, using the coefficients and standard deviations of the explanatory variables, it was found out that the usable living area of the flat has the largest impact on its price from the “non-dummy” variables. In numbers, the change of usable living area of the flat by one standard deviation causes the increase in price by 40.3 % (using the WLS estimate).

## Bibliography

Australian Valuations (2015), *Market value assessments & premise of value*. Available from: < http://australianvaluations.com.au/asset-services/market-value/>. [23 July 2015]

CMS Legal Services EEIG (2011), *CMS Guide to Real Estate Transaction Costs in Europe*. Available from: <http://www.cms-cmck.com/Hubbard.FileSystem/files/Publication/052edf4b-8c7b-40a4-86e5-dfd232d7d389/Presentation/PublicationAttachment/5596207a-5664-4066-b866-e03ac8e29573/CMS%20RE%20Guide%20Transaction%20Cost%20EU%20-%20September%202011.pdf>. [23 July 2015]

COLWELL, P. F. & G. DILMORE (1999): "Who was first? An examination of an early hedonic study." *Land Economics* 75/4: pp. 620-626

Czech National Bank (2014), *Databáze časových řad ARAD*. Available from: < http://www.cnb.cz/cnb/STAT.ARADY_PKG.PARAMETRY_SESTAVY?p_sestuid=19589&p_strid=AAAF&p_lang=CS>. [23 July 2015]

ESHET, T., M. G. BARON, M. SHECHTER & O. AYALON (2007): "Measuring externalities of waste transfer stations in Israel using hedonic pricing." *Waste Management* 27.5: pp. 614-625

ETTER, W. E. (1994): "Direct Capitalization Versus Discounted Cash Flow Analysis.“ *Tierra Grande* 1051, Real Estate Center

FOLLAIN, J. R. & S. MALPEZZI (1980): "Dissecting housing value and rent." Washington, DC: *The Urban Institute*

HILBERS, P., A. W. HOFFMAISTER, A. BANERJI, & H. SHI (2008): "House Price Developments in Europe: A Comparison." *IMF Working Paper* 08/211, International Monetary Fund

HOPKINS, R. E. & M. J. ACTON (1999): "Where does the return come from? Using the risk-adjusted performance measure in real estate." *Real Estate Finance* 16: pp. 23-30

Investment Property Partnership (2015), *Commercial Property Valuation Using The Profits Method*. Available from: <https://investmentproperty.co.uk/property-knowledge-zone/finance-valuation-and-tax/commercial-property-valuation-using-the-profits-method>. [23 July 2015]

LUX, M. (2009): Housing policy and housing finance in the Czech Republic during transition: An example of the schism between the still-living past and the need of reform. Amsterdam: IOS Press. ISBN: 9781607500582

MALPEZZI, S. (2002): “Hedonic Pricing Models: A Selective and Applied Review.” *Section in Housing Economics and Public Policy: Essays in Honour of Duncan Maclennan*

MALPEZZI, S., L. OZANNE & T. THIBODEAU (1980): “Characteristic prices of housing in fifty-nine metropolitan areas.” Washington, DC: *The Urban Institute*

MILES, D. D. (1994): *Housing, financial markets and the wider economy*. Chichester: Wiley. ISBN: 0471952109

MONSON, M. (2009): “Valuation Using Hedonic Pricing Model.” *Cornell Real Estate Review* 7: pp. 62-73

MORANCHO, A. B. (2002): “A hedonic valuation of urban green areas.” *Landscape and Urban Planning* 66.1: pp. 35-41

NAKAMURA, L. (2010): “How much is that home really worth? Appraisal bias and house-price uncertainty.” *Business Review Q1*: pp. 11-22

OTTENSMANN, J. R., S. PAYTON & J. MAN (2008): “Urban locations and housing prices within a hedonic model.” *Journal of Regional Analysis and Policy* 38.1: pp. 19-35

PAGOURTZKI, E., V. ASSIMAKOPOULOS, T. HATZICHRISTOS & N. FRENCH (2003): “Real estate appraisal: A review of valuation methods.“ *Journal of Property Investment & Finance* 21/4: pp. 383-401

PARKER, D. (2011), *The Contractors Method of Commercial Property Valuation*. Available from: <http://www.thepropertyspeculator.co.uk/2011/08/the-contractors-method-of-commercial-property-valuation/>. [23 July 2015]

PropEx PLS (2003), *The Income Capitalization Approach*. Available from: <http://www.propex.com/C_g_inc.htm>. [23 July 2015]

Raiffeisen stavební spořitelna (2013), *Jaké rekonstrukce Češi nejvíc potřebují? Před zimou vede hlavně zateplení*. Available from: < http://www.rsts.cz/jake-rekonstrukce-cesi-nejvic-potrebuji-pred-zimou-vede-hlavne-zatepleni>. [23 July 2015]

SHEPPARD, S. CH. (1998): “Hedonic Analysis of Housing Markets.” *Handbook of regional and urban economics* 3: pp. 1595-1635

SIRMANS, G. S., D. A. MACPHERSON & E. N. ZIETZ (2005): “The Composition of Hedonic Pricing Models.“ *Journal of Real Estate Literature* 13.1: pp. 1-44

Studentskéfinance.cz (2013), *Většina 18 letých požaduje po rodičích vlastní bydlení*. Available from: <http://student.finance.cz/zpravy/finance/394800-vetsina-18-letych-pozaduje-po-rodicich-vlastni-bydleni/>. [23 July 2015]

WOOLDRIDGE, J. M. (2003): *Introductory Econometrics: A Modern Approach*. Cincinnati: Southwestern Thomson Learning .ISBN: 0324113641

# List of appendices

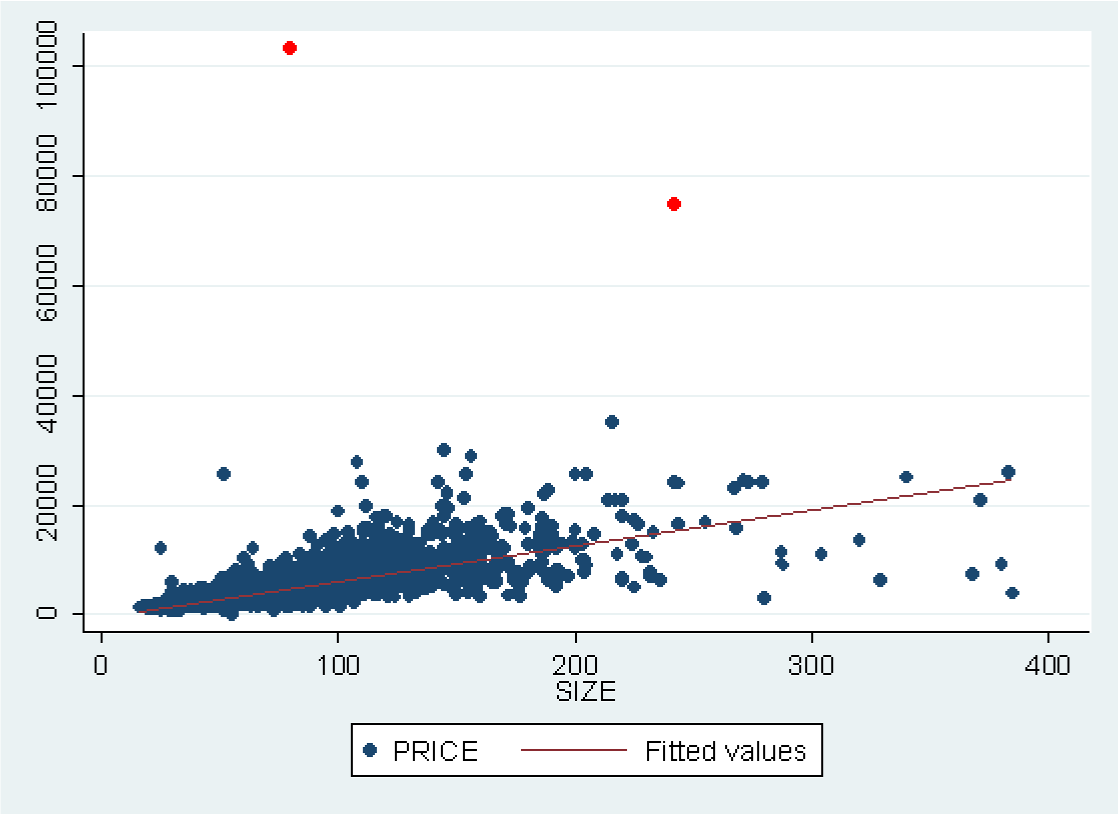
# Appendices

**Appendix 1: Simple summary statistics of the data (table)**

| Variable | Observations | Mean | Standard dev. | Minimum | Maximum |
|---|---|---|---|---|---|
| ATM | 4,604 | 334.6342000 | 214.3386000 | 0 | 1,157 |
| CENDIS | 5,764 | 4,902.3570000 | 3,152.1040000 | 122 | 18,222 |
| ELEVATOR | 4,777 | 0.8392296 | 0.3673575 | 0 | 1 |
| FLOOR | 5,996 | 3.7381590 | 2.3325280 | 0 | 19 |
| FOOD | 4,711 | 191.6090000 | 167.8034000 | 0 | 1,094 |
| KITCHEN | 5,972 | 0.3009042 | 0.4586895 | 0 | 1 |
| mat_BRICK | 5,996 | 0.7001334 | 0.4582375 | 0 | 1 |
| mat_MIXED | 5,996 | 0.0617078 | 0.2406442 | 0 | 1 |
| mat_PANEL | 5,996 | 0.1871247 | 0.3900442 | 0 | 1 |
| mat_SKELETON | 5,996 | 0.0510340 | 0.2200855 | 0 | 1 |
| own_COOPRTV | 5,996 | 0.1195797 | 0.3244965 | 0 | 1 |
| own_PERSONAL | 5,996 | 0.8767512 | 0.3287500 | 0 | 1 |
| own_STATE | 5,996 | 0.0036691 | 0.0604670 | 0 | 1 |
| PHARMACY | 4,461 | 378.3773000 | 248.5528000 | 0 | 1,256 |
| POST | 4,320 | 507.7896000 | 250.0149000 | 0 | 1,277 |
| **PRICE** | **5,913** | **4,654.1450000** | **3,177.7800000** | **220** | **35,000** |
| ROOMS | 5,972 | 2.6364700 | 0.9398041 | 1 | 7 |
| SIZE | 5,996 | 79.9406300 | 38.5755100 | 16 | 410 |
| stt_AFTERREC | 5,996 | 0.2154770 | 0.4111871 | 0 | 1 |
| stt_BEFORREC | 5,996 | 0.0248499 | 0.1556805 | 0 | 1 |
| stt_GOOD | 5,996 | 0.1039026 | 0.3051596 | 0 | 1 |
| stt_NEW | 5,996 | 0.3353903 | 0.4721661 | 0 | 1 |
| stt_POOR | 5,996 | 0.0003336 | 0.0182620 | 0 | 1 |
| stt_UNDERCON | 5,996 | 0.0353569 | 0.1846956 | 0 | 1 |
| stt_VERYGOOD | 5,996 | 0.2846898 | 0.4513042 | 0 | 1 |

Source: www.sreality.cz, STATA

**Appendix 2: Price outliers (graph)**




Source: STATA

**Appendix 3: Housing price determinants (level-level model) (table)**

| Variable | Coefficient | Standard error | P-value | Robust standard error | P-value |
|---|---|---|---|---|---|
| ATM | -0.863345 | 0.1856824 | 0.00 | 0.1774209 | 0.00 |
| CENDIS | -0.264832 | 0.0139664 | 0.00 | 0.0149046 | 0.00 |
| ELEVATOR | 327.4719 | 89.08134 | 0.00 | 75.7493 | 0.00 |
| FLOOR | 28.46768 | 13.7251 | 0.04 | 12.5997 | 0.02 |
| KITCHEN | -293.9176 | 76.51546 | 0.00 | 83.56299 | 0.00 |
| mat_MIXED | -76.27386 | 160.6029 | 0.64 | 195.322 | 0.70 |
| mat_PANEL | -373.5266 | 105.8096 | 0.00 | 84.96643 | 0.00 |
| mat_SKELETON | -82.98655 | 156.7311 | 0.60 | 154.7611 | 0.59 |
| own_COOPRTV | -533.9887 | 104.4356 | 0.00 | 79.92185 | 0.00 |
| own_STATE | 236.6983 | 521.0921 | 0.65 | 240.1834 | 0.32 |
| POST | -0.501263 | 0.1367698 | 0.00 | 0.1353231 | 0.00 |
| ROOMS | 358.8533 | 49.26427 | 0.00 | 89.65718 | 0.00 |
| SIZE | 53.25526 | 1.342827 | 0.00 | 3.336729 | 0.00 |
| stt_AFTERREC | 642.4677 | 114.0689 | 0.00 | 97.41136 | 0.00 |
| stt_BEFORREC | 413.5513 | 201.0487 | 0.04 | 268.8058 | 0.12 |
| stt_NEW | 837.1815 | 124.3892 | 0.00 | 111.5207 | 0.00 |
| stt_POOR | 307.9576 | 1221.075 | 0.80 | 146.5588 | 0.04 |
| stt_UNDERCON | 789.0869 | 245.3 | 0.00 | 283.2013 | 0.01 |
| stt_VERYGOOD | 338.9171 | 110.4699 | 0.00 | 92.08508 | 0.00 |
| C | 559.9943 | 169.8896 | 0.00 | 161.4447 | 0.00 |

n = 3327, R-squared = 0.6916

Source: STATA

**Appendix 4: Housing price determinants (log-level model) (table)**

| Variable | Coefficient | Standard error | P-value | Robust standard error | P-value |
|---|---|---|---|---|---|
| ATM | -7.53E-05 | 0.0000261 | 0.00 | 0.0000258 | 0.00 |
| CENDIS | -4.64E-05 | 1.96E-06 | 0.00 | 2.18E-06 | 0.00 |
| ELEVATOR | 0.0637976 | 0.0124914 | 0.00 | 0.0119764 | 0.00 |
| FLOOR | 0.0049631 | 0.0019164 | 0.01 | 0.0018937 | 0.01 |
| mat_MIXED | 0.0019832 | 0.0225543 | 0.93 | 0.0243805 | 0.94 |
| mat_PANEL | -0.221199 | 0.0145948 | 0.00 | 0.0130607 | 0.00 |
| mat_SKELETON | 0.0124696 | 0.0220128 | 0.57 | 0.0236097 | 0.60 |
| own_COOPRTV | -0.148967 | 0.0146547 | 0.00 | 0.0134182 | 0.00 |
| own_STATE | 0.0667605 | 0.0731876 | 0.36 | 0.0688162 | 0.33 |
| POST | -0.000111 | 0.0000192 | 0.00 | 0.0000192 | 0.00 |
| ROOMS | 0.1529249 | 0.0069194 | 0.00 | 0.0117007 | 0.00 |
| SIZE | 0.0072295 | 0.0001875 | 0.00 | 0.0004181 | 0.00 |
| stt_AFTERREC | 0.1626001 | 0.0158796 | 0.00 | 0.0148636 | 0.00 |
| stt_BEFORREC | 0.0430774 | 0.0281963 | 0.13 | 0.0334283 | 0.20 |
| stt_NEW | 0.2280217 | 0.0168708 | 0.00 | 0.0163114 | 0.00 |
| stt_POOR | 0.1316176 | 0.1714545 | 0.44 | 0.0396049 | 0.00 |
| stt_UNDERCON | 0.2566603 | 0.0341058 | 0.00 | 0.0380272 | 0.00 |
| stt_VERYGOOD | 0.0818121 | 0.015507 | 0.00 | 0.0147002 | 0.00 |
| C | 7.456008 | 0.0232232 | 0.00 | 0.0251742 | 0.00 |

n = 3327, R-squared = 0.7998

Source: STATA

**Appendix 5: Housing price determinants (log-log model) (table)**

| Variable | Coefficient | Standard error | P-value | Robust standard error | P-value |
|---|---|---|---|---|---|
| lCENDIS | -0.214977 | 0.0066484 | 0.00 | 0.0071056 | 0.00 |
| ELEVATOR | 0.0495791 | 0.0111766 | 0.00 | 0.0107191 | 0.00 |
| FLOOR | 0.0066291 | 0.0017061 | 0.00 | 0.0018237 | 0.00 |
| Lfood | 0.0138616 | 0.0050219 | 0.01 | 0.0048347 | 0.00 |
| KITCHEN | -0.044002 | 0.0097328 | 0.00 | 0.0100288 | 0.00 |
| mat_MIXED | 0.0196546 | 0.0199785 | 0.33 | 0.0239234 | 0.41 |
| mat_PANEL | -0.175622 | 0.0131725 | 0.00 | 0.0115514 | 0.00 |
| mat_SKELETON | 0.0162739 | 0.0197002 | 0.41 | 0.0212799 | 0.44 |
| own_COOPRTV | -0.151003 | 0.0131362 | 0.00 | 0.0131782 | 0.00 |
| own_STATE | 0.0463383 | 0.0650402 | 0.48 | 0.0550491 | 0.40 |
| lPOST | -0.033674 | 0.0062417 | 0.00 | 0.0067796 | 0.00 |
| ROOMS | 0.0872014 | 0.0071113 | 0.00 | 0.0089016 | 0.00 |
| lSIZE | 0.7308737 | 0.0166449 | 0.00 | 0.0209647 | 0.00 |
| stt_AFTERREC | 0.1539572 | 0.014293 | 0.00 | 0.0133308 | 0.00 |
| stt_BEFORREC | 0.0164007 | 0.0258032 | 0.53 | 0.0314928 | 0.60 |
| stt_NEW | 0.2049929 | 0.0155579 | 0.00 | 0.0144228 | 0.00 |
| stt_POOR | 0.1860191 | 0.1525599 | 0.22 | 0.0135554 | 0.00 |
| stt_UNDERCON | 0.2273262 | 0.0307148 | 0.00 | 0.0377352 | 0.00 |
| stt_VERYGOOD | 0.0730981 | 0.0138197 | 0.00 | 0.0129508 | 0.00 |
| C | 6.69459 | 0.0849575 | 0.00 | 0.0961781 | 0.00 |

n = 3281, R-squared = 0.8369

Source: STATA

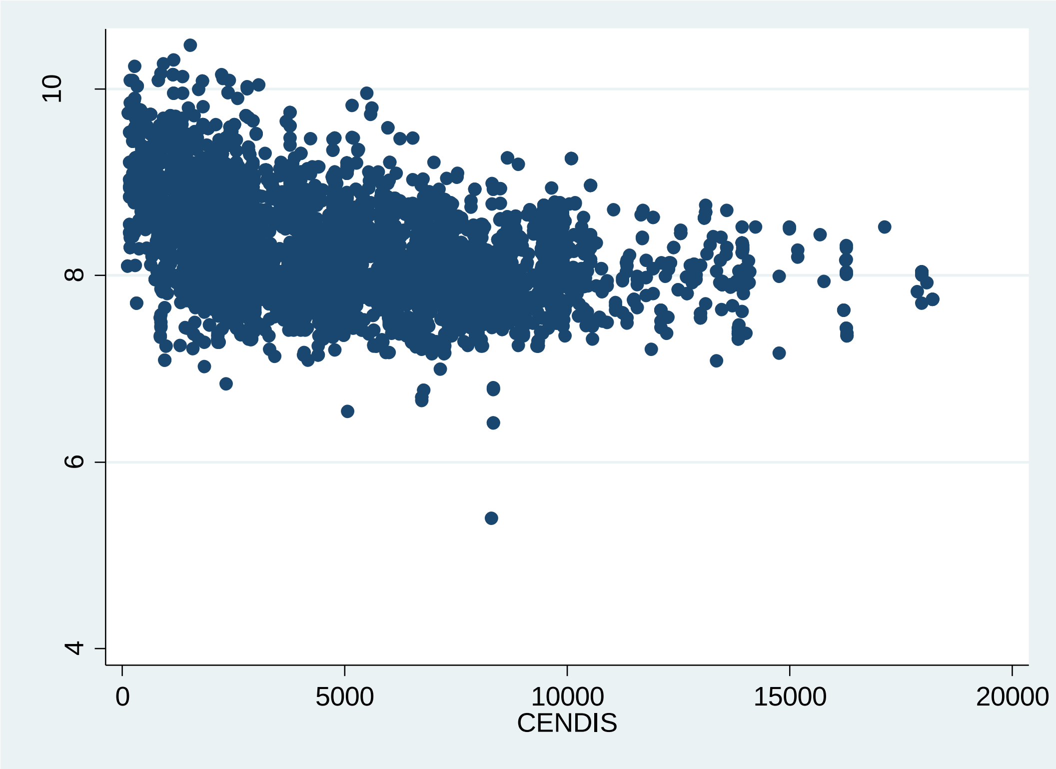
**Appendix 6: Log-level model with CENDIS2 (table)**

| Variable | Coefficient | Robust standard error | P-value |
|---|---|---|---|
| ATM | -0.0000252 | 0.0000284 | 0.38 |
| **CENDIS** | **-0.0001181** | **0.00000635** | **0.00** |
| **CENDIS2** | **5.84E-09** | **4.69E-10** | **0.00** |
| ELEVATOR | 0.0591214 | 0.0118004 | 0.00 |
| FLOOR | 0.00473 | 0.0018894 | 0.01 |
| FOOD | 0.000062 | 0.0000348 | 0.08 |
| KITCHEN | -0.0112599 | 0.011242 | 0.32 |
| mat_MIXED | 0.0262031 | 0.0241358 | 0.28 |
| mat_PANEL | -0.1774843 | 0.0129444 | 0.00 |
| mat_SKELETON | 0.0332488 | 0.0250114 | 0.18 |
| own_COOPRTV | -0.1482856 | 0.0132691 | 0.00 |
| own_STATE | 0.0504936 | 0.150095 | 0.74 |
| PHARMACY | -3.24E-06 | 0.0000243 | 0.89 |
| POST | -0.0001061 | 0.0000203 | 0.00 |
| ROOMS | 0.1546573 | 0.0112657 | 0.00 |
| SIZE | 0.0070256 | 0.0004036 | 0.00 |
| stt_AFTERREC | 0.1585655 | 0.0148591 | 0.00 |
| stt_BEFORREC | 0.0373714 | 0.0320215 | 0.24 |
| stt_NEW | 0.2433228 | 0.0170481 | 0.00 |
| stt_POOR | 0.1813242 | 0.0188553 | 0.00 |
| stt_UNDERCON | 0.2761266 | 0.0370321 | 0.00 |
| stt_VERYGOOD | 0.0783108 | 0.0145152 | 0.00 |
| C | 7.591728 | 0.0286757 | 0.00 |

n = 3248, R-squared = 0.8109

Source: STATA

**Appendix 7: Relationship between CENDIS and lPRICE (graph)**




Source: STATA

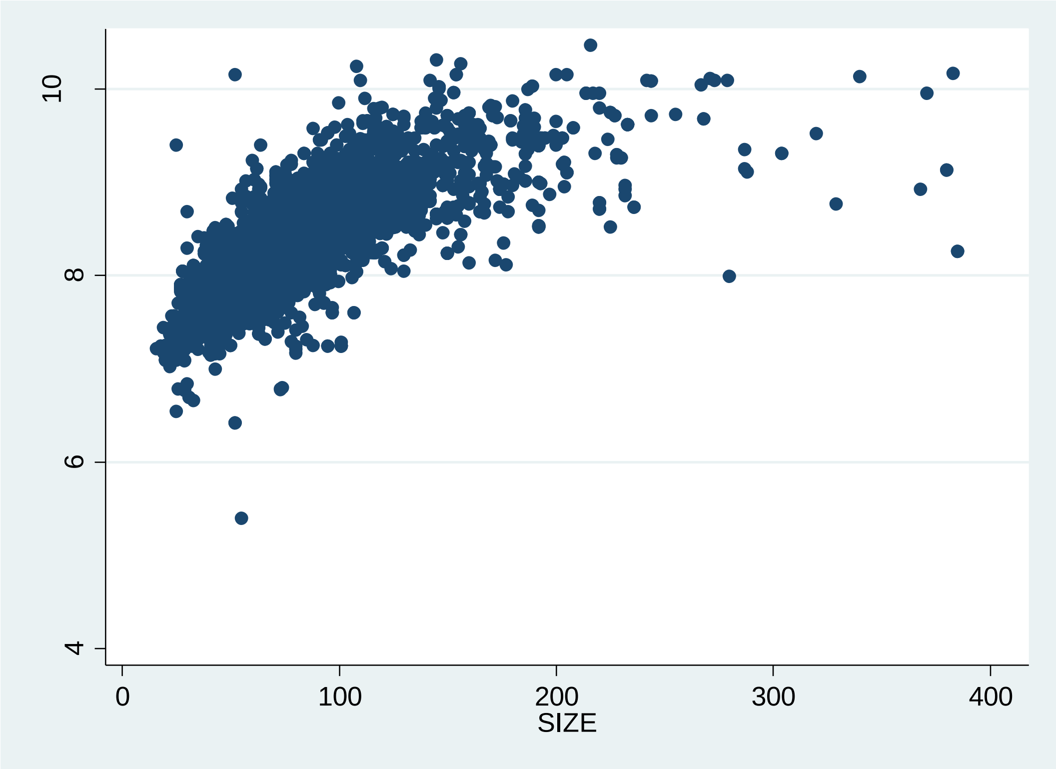
**Appendix 8: Log-level model with SIZE2 (table)**

| Variable | Coefficient | Robust standard error | P-value |
|---|---|---|---|
| ATM | -0.0001161 | 0.0000276 | 0.00 |
| CENDIS | -0.0000475 | 0.00000203 | 0.00 |
| ELEVATOR | 0.0585583 | 0.011566 | 0.00 |
| FLOOR | 0.005231 | 0.0018294 | 0.00 |
| FOOD | 0.0001046 | 0.000032 | 0.00 |
| KITCHEN | -0.0380338 | 0.010738 | 0.00 |
| mat_MIXED | -0.0135883 | 0.0245336 | 0.58 |
| mat_PANEL | -0.1816972 | 0.0128115 | 0.00 |
| mat_SKELETON | -0.0043966 | 0.0229446 | 0.85 |
| own_COOPRTV | -0.1499152 | 0.0133111 | 0.00 |
| own_STATE | 0.0711049 | 0.1328476 | 0.59 |
| PHARMACY | -0.0000108 | 0.0000246 | 0.66 |
| POST | -0.0000941 | 0.0000193 | 0.00 |
| ROOMS | 0.0803379 | 0.0100509 | 0.00 |
| **SIZE** | **0.0160393** | **0.0008384** | **0.00** |
| **SIZE2** | **-0.0000322** | **0.00000334** | **0.00** |
| stt_AFTERREC | 0.1526478 | 0.0145484 | 0.00 |
| stt_BEFORREC | 0.0327692 | 0.0320588 | 0.31 |
| stt_NEW | 0.2064676 | 0.0163664 | 0.00 |
| stt_POOR | 0.1621035 | 0.0217141 | 0.00 |
| stt_UNDERCON | 0.2125041 | 0.0401008 | 0.00 |
| stt_VERYGOOD | 0.0818246 | 0.0141282 | 0.00 |
| C | 7.206025 | 0.0338383 | 0.00 |

n = 3248, R-squared = 0.8251

Source: STATA

**Appendix 9: Relationship between lPRICE and SIZE (graph)**




Source: STATA

**Appendix 10: Log-level model with ELEVFLOOR (table)**

| Variable | Coefficient | Robust standard error | P-value |
|---|---|---|---|
| ATM | -0.0000823 | 0.0000289 | 0.00 |
| CENDIS | -0.0000489 | 0.00000224 | 0.00 |
| ELEVATOR | 0.0729021 | 0.025183 | 0.00 |
| **FLOOR** | **0.0081844** | **0.0060238** | **0.17** |
| **ELEVFLOOR** | **-0.0034641** | **0.006326** | **0.58** |
| FOOD | 0.0000675 | 0.0000355 | 0.06 |
| KITCHEN | -0.0137953 | 0.0115799 | 0.23 |
| mat_MIXED | 0.0013662 | 0.0246654 | 0.96 |
| mat_PANEL | -0.2124026 | 0.0132235 | 0.00 |
| mat_SKELETON | 0.0051608 | 0.0246389 | 0.83 |
| own_COOPRTV | -0.1498616 | 0.0134139 | 0.00 |
| own_STATE | 0.073116 | 0.1473613 | 0.62 |
| PHARMACY | -0.0000175 | 0.0000249 | 0.48 |
| POST | -0.0001159 | 0.0000206 | 0.00 |
| ROOMS | 0.1544036 | 0.0118923 | 0.00 |
| SIZE | 0.0072239 | 0.0004256 | 0.00 |
| stt_AFTERREC | 0.1609553 | 0.0152105 | 0.00 |
| stt_BEFORREC | 0.0489316 | 0.0335328 | 0.15 |
| stt_NEW | 0.2257973 | 0.0173424 | 0.00 |
| stt_POOR | 0.135581 | 0.0327416 | 0.00 |
| stt_UNDERCON | 0.2586634 | 0.0380116 | 0.00 |
| stt_VERYGOOD | 0.0822534 | 0.0148168 | 0.00 |
| C | 7.456908 | 0.0321714 | 0.00 |

n = 3248, R-squared = 0.8010

Source: STATA

**Appendix 11: Comparison of OLS with WLS (table)**

| | **Coefficient** | | **Standard error** | | **P-value** | |
|---|---|---|---|---|---|---|
| **Variable** | **OLS** | **WLS** | **OLS (robust)** | **WLS** | **OLS** | **WLS** |
| ATM | -5.82E-05 | -6.32E-05 | 0.0000241 | 0.00002 | 0.02 | 0.00 |
| CENDIS | -0.000108 | -8.92E-05 | 0.00000525 | 4.75E-06 | 0.00 | 0.00 |
| CENDIS2 | 5.09E-09 | 3.72E-09 | 3.81E-10 | 3.47E-10 | 0.00 | 0.00 |
| ELEVATOR | 0.0574532 | 0.0329806 | 0.0111244 | 0.0091985 | 0.00 | 0.00 |
| FLOOR | 0.0049758 | 0.0037695 | 0.0017952 | 0.0014228 | 0.01 | 0.01 |
| FOOD | 0.0000944 | 0.0000879 | 0.0000271 | 0.0000215 | 0.00 | 0.00 |
| KITCHEN | -0.035407 | -0.05443 | 0.0104382 | 0.0090621 | 0.00 | 0.00 |
| mat_MIXED | 0.0138129 | 0.0466044 | 0.0237637 | 0.0164347 | 0.56 | 0.01 |
| mat_PANEL | -0.152956 | -0.155665 | 0.0124152 | 0.0097654 | 0.00 | 0.00 |
| mat_SKELETON | 0.0206168 | 0.0079335 | 0.0214832 | 0.0176229 | 0.34 | 0.65 |
| own_COOPRTV | -0.150105 | -0.131797 | 0.0131161 | 0.0111263 | 0.00 | 0.00 |
| own_STATE | 0.0651525 | 0.0586455 | 0.0626945 | 0.0381739 | 0.30 | 0.13 |
| POST | -0.000084 | -7.19E-05 | 0.0000184 | 0.0000131 | 0.00 | 0.00 |
| ROOMS | 0.0829842 | 0.083859 | 0.0096926 | 0.0064628 | 0.00 | 0.00 |
| SIZE | 0.0154997 | 0.0159675 | 0.0007928 | 0.0005573 | 0.00 | 0.00 |
| SIZE2 | -3.09E-05 | -3.45E-05 | 0.00000314 | 2.42E-06 | 0.00 | 0.00 |
| stt_AFTERREC | 0.1492343 | 0.1199248 | 0.0142272 | 0.0128558 | 0.00 | 0.00 |
| stt_BEFORREC | 0.0214782 | -0.037817 | 0.0307725 | 0.022016 | 0.49 | 0.09 |
| stt_NEW | 0.2172062 | 0.1934515 | 0.0160566 | 0.0139129 | 0.00 | 0.00 |
| stt_POOR | 0.1988261 | 0.1740601 | 0.0135369 | 0.0123465 | 0.00 | 0.00 |
| stt_UNDERCON | 0.2235757 | 0.2228359 | 0.0390474 | 0.0372686 | 0.00 | 0.00 |
| stt_VERYGOOD | 0.0753822 | 0.0556673 | 0.0138692 | 0.0121013 | 0.00 | 0.00 |
| C | 7.329285 | 7.312804 | 0.0339107 | 0.0243223 | 0.00 | 0.00 |

n = 3327, R-squared = 0.8327 (OLS); 0.8756 (WLS)

Source: STATA