CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Regional Analysis of Real Estate Bubbles in the Czech Republic

Bachelor's thesis

Author: Dominik Benk

Study program: Economics and Finance

Supervisor: Mgr. Petr Polák MSc.

Year of defense: 2019

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, July 28, 2019

Dominik Benk

Abstract

The recent increase in real estate prices in the Czech Republic had a serious impact on the accessibility of housing. This thesis aimed to assess, whether it was justified by changes in economic fundamentals via the approach of inverted demand. Two distinct models were described and each was applied to a separate dataset. The crucial difference is the way these models deal with non-stationary data, i.e. either the variables are differenced or the cointegrating relationship is examined. Estimation itself was done by Fixed-effects or Dynamic OLS, and the latter model additionally allowed to evaluate the speed of adjustment to equilibrium. Various macroeconomic factors were considered out of which wages, unemployment, natural increase in population, net migration, construction prices and real interest rates turned out to be fundamental drivers of flat prices. Real estate price bubbles were then evaluated with Price to income and Price to rent indicators, followed by an analysis of differences between actual and fundamental prices, which utilized models of determinants. The results suggested 2 main periods of overvaluation, namely 2003 and 2008, and a high risk of current overvaluation in Prague and region Jihomoravský.

JEL Classification	C33, C51, R21, R31
Keywords	real estate prices, price bubbles, fixed effects, error correction model, regional data, panel study, Czech Republic

Title	Regional Analysis of Real Estate Bubbles in the Czech Republic

Author's e-mail	`benk.dominik@seznam.cz`
Supervisor's e-mail	`polakpet@gmail.com`

Abstrakt

Nedávný nárůst cen nemovitostí v České republice měl vážný dopad na dostupnost bydlení. Cílem práce bylo posoudit, zda to bylo odůvodněno změnami ekonomických fundamentů a to prostřednictvím přístupu invertované poptávky. Byly popsány dva odlišné modely a každý byl aplikován na odlišná data. Hlavní rozdíl je v tom, jak se tyto modely vypořádávají s nestacionárními daty, tj. buď jsou proměnné diferencovány, nebo je zkoumán kointegrační vztah. Samotný výpočet byl proveden pomocí Fixních efektů nebo Dynamického OLS a druhý model navíc umožnil vyhodnotit rychlost návratu k rovnovážnému stavu. Byly zváženy různé makroekonomické faktory, z nichž se mzdy, nezaměstnanost, přirozený nárůst obyvatelstva, čistá migrace, ceny staveb a reálné úrokové sazby ukázaly být fundamentálními hnacími silami cen bytů. Realitní cenové bubliny pak byly vyhodnoceny pomocí indikátorů Price to income a Price to rent, následovaných analýzou rozdílů mezi skutečnými a fundamentálními cenami, využívající modely determinantů. Výsledky naznačily 2 hlavní období nadhodnocení, konkrétně 2003 a 2008, a vysoké riziko současného nadhodnocení v Praze a Jihomoravském kraji.

Klasifikace JEL	C33, C51, R21, R31
Klíčová slova	ceny nemovitostí, cenové bubliny, fixní efekty, model korekce chyb, regionální data, panelová studie, Česká republika

Název práce	Regionální Analýza Realitních Bublin v České Republice

E-mail autora	`benk.dominik@seznam.cz`
E-mail vedoucího práce	`polakpet@gmail.com`

Acknowledgments

I would like to express my gratitude to Mgr. Petr Polák MSc. for providing me with valuable feedback and useful comments. Special thanks also to my family and friends for their support and motivation.

Typeset in FSV LaTeX template with great thanks to prof. Zuzana Havránková and prof. Tomáš Havránek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

Bibliographic Record

Benk, Dominik: *Regional Analysis of Real Estate Bubbles in the Czech Republic.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2019, pages 64. Advisor: Mgr. Petr Polák MSc.

Contents

List of Tables

List of Figures

Acronyms

CZSO Czech Statistical Office

CNB Czech National Bank

IRI Institute for Regional Information

CEE Central and Eastern Europe

OECD Organization for Economic Cooperation and Development

POLS Pooled ordinary least squares

FE Fixed-Effects

DOLS Dynamic OLS

FMOLS Fully modified OLS

ECM Error correction model

ECT Error correction term

VAR Vector autoregression

VECM Vector error correction model

GSADF Generalized sup Augmented Dickey-Fuller test

CLM Classical linear model

ADF Augmented Dickey-Fuller

CADF Cross-sectionally augmented Dickey-Fuller

PP Phillips-Perron

LLC Levin, Lin, and Chu

IPS Im, Pesaran and Shin

CIPS Cross-sectionally augmented Im, Pesaran and Shin

HP Hodrick-Prescott

P/I Price-to-income

R/I Price-to-rent

Bachelor's Thesis Proposal

Author	Dominik Benk
Supervisor	Mgr. Petr Polák MSc.
Proposed topic	Regional Analysis of Real Estate Bubbles in the Czech Republic

Research question and motivation

The purpose of the thesis is to analyze prices in the real estate market from a perspective that slightly differs from the usual point of view. In hedonic models our estimations are dependent on the characteristics of the dwelling (e.g. area, location, floor number) and the explained variable is the observed price (or price offered). I will focus on the factors that influence demand (such as income, mortgage rate, number of households) instead, even though it might be challenging. In the practical part, the thesis will focus on the regions of the Czech Republic, primarily evaluating the diversity of the real estate market in Prague in comparison with the other regions.

Contribution

The objective is to evaluate what the observed differences between Prague and the rest of the Czech Republic are, and whether they can be rationally explained from the perspective of demand. Also, in reaction to CNB considering real estate prices overvalued, I will analyze the recent trends to see whether our model suggests the same. In particular I would like to assess the adequacy of an increase in prices to an increase in income and make a conclusion about the possibility of the presence of a price bubble. To my knowledge there was a similar study performed by Michal Hlaváček and Luboš Komárek, who analyzed prices in the CR during the period of 1998-2008. Since then there has been new data published and I consider it a great opportunity to examine the recent surge in prices.

Last but not least, I will also estimate the income elasticity of demand for real-estate and its dynamics during the recent years. This could find use in understanding the impacts of imposing many public policies.

Methodology

Using multiple linear regression and panel data, the model will represent the relationship between the average prices of residential real estate and fundamentals such as income of individuals, unemployment, several demographic factors, (number of households, divorces, marriages, average age) and mortgage interest rate. I will also consider the return on investment, tenure choice, and short-term rentals (Airbnb for example).

According to data, I will focus separately on flats and family houses in each of the regions of the Czech Republic that are monitored by the Czech Statistical Office. Apart from that, important data sources are the Ministry for Regional Development, the Czech National Bank and Eurostat.

Outline

1. Introduction
2. Literature review
3. Methodology
4. Data analysis
5. Conclusion

Core bibliography

1. Hlaváček, M., and Komárek, L. (2010). Rovnovážnost cen nemovitostí v České republice. Politická ekonomie, 58(3), 326-342.
2. Égert, B., and Mihaljek, D. (2007). Determinants of house prices in central and eastern Europe. Comparative economic studies, 49(3), 367-388.
3. Mulford, J. E. (1979). Income Elasticity of Housing Demand. Rand Corporation.
4. Gallin, J. (2006). The long-run relationship between house prices and income: evidence from local housing markets. Real Estate Economics, 34(3), 417-438.

5. ČSÚ (2017). Ceny sledovaných druhů nemovitostí - 2014 až 2016. Available at: `https://www.czso.cz/csu/czso/`

6. Green, R., and Hendershott, P. H. (1996). Age, housing demand, and real house prices. Regional Science and Urban Economics, 26(5), 465-480. De Leeuw, F. (1971). The demand for housing: A review of cross-section evidence. The Review of Economics Statistic, 1-10.

7. Attanasio, O. P., Bottazzi, R., Low, H. W., Nesheim, L., and Wakefield, M. (2012). Modelling the demand for housing over the life cycle. Review of Economic Dynamics, 15(1), 1-18.

8. Manrique, J., and Ojah, K. (2003). The demand for housing in Spain: an endogenous switching regression analysis. Applied Economics, 35(3), 323-336.

9. Garcia, J., and Raya, J. M. (2011). Price and income elasticities of demand for housing characteristics in the city of Barcelona. Regional Studies, 45(5), 597-608.

10. Tsai, I. C. (2013). Housing affordability, self-occupancy housing demand and housing price dynamics. Habitat International, 40, 73-81.

11. Gary-Bobo, R. J., and Larribeau, S. (2003). The bank's market power and the interest-rate elasticity of demand for housing: an econometric study of discrimination on French mortgage data. Annales d'Economie et de Statistique, 377-398.

Author   Supervisor

Chapter 1

Introduction

During recent years, the real estate market has earned considerable attention from economists as well as the general public. The surge in prices has raised concerns about threats similar to the subprime mortgage crisis in 2007. The evolution of prices seems to be supported by significant economic growth, low unemployment, increasing real wages and optimism about future developments. Also, thanks to the current trend of owning a flat as an attractive investment, insufficient supply appears to be overwhelmed by demand. On the other hand, since 2017, Czech national bank (CNB) is following a policy of continuous increases in interest rates, indirectly affecting mortgage rates and lowering the affordability of housing. Nevertheless, there is no clear indication of potential overheating of the market and a sudden decline in the future.

It is no surprise that real estate markets play a vital role in every national economy. As they are closely linked to various areas, housing prices that are not aligned with development in fundamentals represent a substantial threat to financial stability and may be the initial trigger for a recession. Being able to determine what drives the prices and to what extent they correspond to economic fundamentals is without doubt a very powerful tool. It allows not only to estimate the real fundamental value and predict future dynamics, but more importantly to avoid or at least mitigate impacts of bursting price bubbles.

Different methods to assess the determinants of real estate prices will be applied, and then the possibility of the presence of a bubble will be evaluated. Various approaches exist on how this can be achieved, depending on different assumptions and level of sophistication. The thesis will draw from selected literature that was considered the most relevant, taking into account the character of the real estate market in the Czech Republic, especially regional differences. Alternatively to hedonic models, in which estimations depend on the characteristics of the dwelling, I will focus on fundamentals such as income of individuals, unemployment, demographic factors and mortgage interest rate. Dynamics of real estate market will be captured by flat prices and in contrast to similar studies that gather offered prices from reality agencies, I will work with realized prices monitored by the Czech Statistical Office (CZSO).

The thesis aims to provide answers to questions such as “What are the fundamentals that drive real estate prices in the Czech Republic?”,“Do we currently observe misalignments in housing prices and what is their severity?”,“Which were the periods of undervaluation and overvaluation in history?”,“Do regional real estate markets respond differently to changes in underlying fundamentals?”. Findings could contribute to an increase in the effectiveness of macroeconomic policies for maintaining financial stability, for instance improve the stress testing framework conducted by CNB.

The rest of the thesis is organized as follows. Chapter 2 reviews literature related to this study. Chapter 3 describes the available data for real estate prices as well as its possible determinants. In Chapter 4, the methodology applied for assessing house price determinants will be explained and basic characteristics of required statistical tests will be provided. Chapter 5 presents the results, evaluates real estate bubbles and discusses possible shortcomings of implemented models. Finally, Chapter 6 summarizes the main findings.

Chapter 2

Literature Review

Literature regarding the topic of housing price determinants, housing bubbles or real estate in general is very rich. However, studies focusing on the domestic market are limited, and therefore I opted to also describe works from foreign authors. It has been shown that various approaches can be applied, and the most relevant for this thesis will be reviewed.

Hlaváček and Komárek (2009) were the first who applied econometric methods to real estate prices in the Czech Republic while considering regional differences. To determine the periods of house price bubbles, three main approaches were used. In particular, simple ratios of price to income and price to rent, followed by time series analysis of aggregate data and panel analysis for regions. The first case of empirical analysis employed quarterly data for house price growth and compared estimated equilibrium prices to Hodrick-Prescott (HP) filter. In the latter case, annual data for price levels were used, models of Pooled ordinary least squares (POLS) and Fixed effects (FE) with constant slope were estimated and residual analysis was performed for all regions. According to findings, most of the demand factors play a vital role in explaining variations in house prices, while supply factors were found to be significantly less important. Property price bubbles were identified in two periods, 2002/2003 and 2007/2008, regardless of methods used. However, results from price sustainability indicators and panel regression were not consistent for the Prague region, signalling that “Prague is specific in nature compared to the other Czech regions”. Authors also admit that mainly because of the short length of the time series used, they were not able to account for problems with endogeneity of explanatory variables, as well as the fact that the included bubble periods may have a negative impact on the estimated price equilibrium.

A similar set of supply and demand factors was used also for example by Belke and Keil (2018), who focused solely on identification of fundamentals and conducted a regional analysis of house and apartment prices in nearly 100 German cities. POLS model was employed as a benchmark for one-way as well as two-way FE, and the results confirmed the dependence of real estate prices not only on the price dynamics itself, but also on underlying drivers. Robustness was then checked via the implementation of different modifications, such as inclusion of lag of dependent variable, substitution of real series by nominal or alternating included variables.

A close link between rents and housing prices was studied by Mikhed and Zemčík (2009), who described a structural model of the housing market, in which they interconnected concepts of present value and inverted demand, based on an assumption of a consumer being indifferent between renting and owning a house. The main purpose of the study was to investigate price bubbles in the U.S. housing market. They emphasized the consideration of multiple fundamental factors, in contrast to similar studies where only one was used. To examine house price dynamics and evaluate the possibility of the presence of a price bubble, a two-stage procedure developed by Engle and Granger (1987) was employed. Zemčík (2011) in his more recent work also evaluated the relationship between Czech real estate prices and rents by applying similar panel data stationarity and causality techniques. He conducted tests for regions of the Czech Republic and districts of Prague separately. The former was based on annual data from the Institute of Regional Information and the latter on monthly data from real estate magazine REALIT. Zemčík came to the conclusion that real estate in the Czech Republic overall as well as in Prague was slightly overvalued, even though the degree of overpricing was inferior to the price bubble in the United States.

A different approach was applied by Égert and Mihaljek (2007), who studied fundamental determinants of house prices in CEE countries in contrast to OECD. The technique of combining Panel dynamic OLS (DOLS) with Error correction model (ECM) was employed. The advantages of such setup lie in implementation of cross-country heterogeneity in both cases of long run and short run together with an allowance for endogeneity in explanatory variables and autocorrelation in residuals. In addition to conventional fundamentals, they also considered various demographics, labour, supply-side and transition specific factors and in order to test for stationarity of series, four different panel unit root tests were performed. The study confirmed findings from empirical literature as the price elasticities with respect to fundamentals as well as the rate of convergence of prices to equilibrium tended to be higher for smaller or catching up economies. On the other hand, possible shortcomings of the results due to lack of available data were emphasized, as there was no option to estimate the model for small OECD countries, and use it as a benchmark and correct for problems with excessive growth of prices in CEE.

Housing price bubbles in the Czech Republic were also examined by Čadil (2009). The author evaluated a simple price-to-income ratio, followed by a more advanced technique of Vector autoregression (VAR). He also provided a comparison of two types of real estate market, namely flats and family houses, and introduced formula to calculate "average share of factor on overall price dynamics", allowing him to evaluate the importance of significant variables. Furthermore, the relationship between mortgages and real estate prices was analysed via Granger causality, showing the possible indication of "speculative motives" in case of flats. In the regression model, differentiated logarithms of variables were used and speculative demand was incorporated via the inclusion of a price lag. Its results suggested significant demand orientation of real estate prices in the Czech Republic. Nevertheless, the author acknowledged substantial regional differences that could not be taken into account because of the absence of data together with short time series (1998-2006), and admitted that "conclusions of this analysis cannot be treated as fully reliable".

A more advanced approach was applied to Kenyan real estate market by Njoroge et al. (2018). The study was based on standard tests for stationarity, cointegration and Granger causality and deepened the analysis through an alternative version of the Dickey-Fuller test. House prices were found not to be Granger caused by macroeconomic variables, and because of no cointegration between nonstationary variables, the Vector error correction model (VECM) could not be employed. The bubble was further examined by Generalized Sup Augmented Dickey-Fuller test (GSADF) on price to rent ratio, evaluating the explosiveness in the time series. The size of the bubble was then measured by a "bubble term", obtained from estimating OLS regression.

Due to the importance of real estate markets for financial stability, it is also possible to draw from approaches that are applied by central banks. Hlaváček and Hejlová (2014) provided a brief overview of methodology currently employed by CNB. [1] Authors emphasized the importance of using multiple approaches simultaneously, each granting additional information, that are later combined into an aggregated model of apartment price gap.[2] The following two econometric and two statistical models were described:

General supply and demand model - Simple linear regression model of many independent variables without any lags. House price misalignments are represented by the portion that is not explained by determinants. Stationarity in series is achieved by taking logarithms or differencing.

Accelerator model - Considers aggregate economic relationships between "the business and credit cycle and the house price cycle" by employing VECM. Cycles are represented by GDP, housing loans and house price index, respectively. The model allows to address endogeneity in variables. The gap in apartment prices is then given in a similar way as in a previous approach.

Economic sense of home ownership - Overvaluations are analysed from the perspective of investment demand, i.e. seeing residential property as future cash inflow. CNB adjusts the ordinary price to rent ratio for "mortgage servicing costs net of tax deductions for mortgage interest".[3] Evaluation then lies in examining deviations of this statistics from an equilibrium value, which is estimated using HP filter.

Affordability of housing - Analysis of consumption demand via the ratio of price-to-income. Similarly to the previous method, this ratio is also adjusted and evaluated with HP filter.

As problems may occur with interpretation due to variance of results from individual approaches, aggregation procedure is then done by deriving two sets of weights from the correlation matrix of price gap estimates given by each approach. The first set gives greater weights to approaches more correlated with others, indicating "mutual nearness", and vice versa for the second set,

[1] CNB results published regularly in Financial Stability Reports.

[2] Dynamics in housing prices are captured solely by apartment prices, because of significant level of heterogeneity of family houses.

[3] Based on standardized 20 years mortgage with 65% loan to value.

indicating “mutual farness”. Because of the inaccuracy of estimation methods, band representing 95% confidence interval is also provided, allowing to consider significance of price deviations. Possible endpoint bias that may arise due to factors connected with the convergence of economy, is assessed via the application of approaches to different time samples separately. Authors also stressed the complexity of the real estate market and the importance of taking into account all possible drawbacks of CNB procedure and that the results should be only treated as a “guide”.

Possible shortcomings of econometric analyses were discussed for instance by Sunega et al. (2014). In particular, they presented two ECM models of house price determinants and showed how they can fail to forecast actual price misalignments in the market. The main difference was in the specification, i.e. the inclusion of interest rate as one of the fundamentals explaining long term volatility in prices. Both models satisfied all crucial assumptions, had strong explanatory power and used the same data for the Czech Republic, however, while the first indicated undervaluation, predictions of the second were close to equilibrium. Authors stressed the need for modification of standard econometric approaches due to the complexity of housing demand as they claimed that it is also affected by “social constructs, social norms, ideologies, unrealistic expectations, symbolic patterns, and that the actual choice of housing is the outcome of complex social interactions with reference groups”.

Alternatively to all these “conventional” approaches described above, one distinct method can also be mentioned, even though it is less related to this thesis. Thanks to development in artificial intelligence, methods such as machine learning found use in many different fields and are also becoming popular even in the case of real estate market modelling. For example, Ćetković et al. (2018) trained an artificial neural network model to forecast real estate prices in 27 EU countries. The predictions were based on 11 input variables representing macroeconomic factors and the impact on price was evaluated for each variable separately. The main advantage of such setup is the fact that it does not restrict the relationships in variables to be linear which is often assumed in standard econometric approaches.

Chapter 3

Data

This section will focus on the process of collecting data and its description. Overall it was quite a complicated procedure, as in a lot of cases, it was necessary to collect the data for each region separately. Nevertheless, I managed to collect 2 balanced panel data sets (quarterly and annual) for 14 regions in the Czech Republic covering the period from 2000 to 2017. The dependent variable will be described in more detail than the explanatory, since to my knowledge, there have been very few papers employing this exact same series.[1]

3.1 Real Estate Prices

Various sources publish data on real estate prices, but the time series usually differ significantly. This is mainly because of a high degree of heterogeneity of the real estate market, and it is essential to choose the "correct" one for analysis, as this might otherwise lead to different results. A brief overview of the sources of real estate data for the Czech Republic is given for example in Hlaváček and Komárek (2009). The first main category the author described were supply prices, which are published for example by the Institute for Regional Information (IRI), who introduced the concept of "standard apartment". While this series allows to control for structural changes, it provides little information about the market as a whole. Other sources include various journals (Realit, King Sturge) or real estate agencies themselves (Lexxus). Generally, they work with bid prices collected from real estate agencies, which might have both pros and cons. Nevertheless, access to their databases is very limited.

[1] Yearly data (1998-2008) employed by Hlaváček and Komárek (2009); Quarterly data (2005-2013) employed by Cempírek (2014).

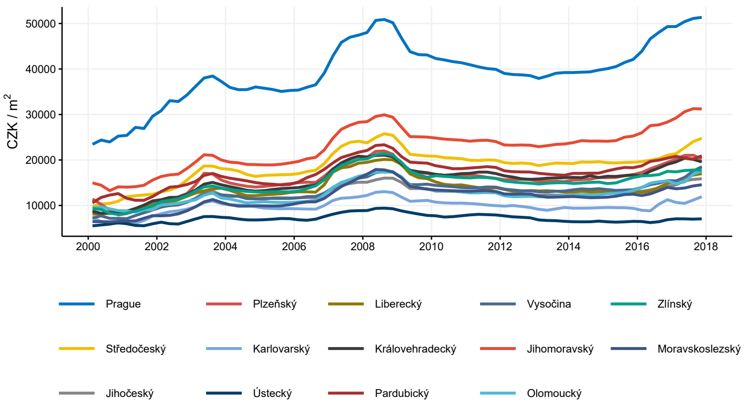
The second group consists of transfer prices. These are available thanks to cooperation of Ministry of Finance with Czech Statistical Office in monitoring real estate prices in the Czech Republic. The main advantage of this data set lies in its accuracy, which comes from the fact that CZSO is provided with information on actual transfer prices admitted in the statement for stamp duty land tax. They are also very appealing because of their completeness and regularity in their reporting, while also accounting for regional differences. Possible shortcomings might be that the admitted price is not completely in line with real transfer price, caused by possible tax optimization. On the other hand, these concerns are not so serious, since if it comes to relative comparison during some period, we would expect this misalignment to remain approximately the same over time. Another disadvantage is the long delay in collecting and publishing the final data, which is reflected in the ending year of series. Also, prices of newly built dwellings are not included as they are not subject to transfer tax. Despite these weaknesses, transfer prices published by CZSO appear to be the most appropriate for my analysis. This dataset is also very complex and needs to be discussed in more detail.

Since 2003, CZSO publishes data on a yearly basis and each publication comes with information for 3 consecutive years. Even though individual publications overlap, price estimates differ among them. This is because of the fact, that they are recalculated over time and the most accurate are the estimates for the first years.[2] Each of these publications provides information for family houses, flats, multi-dwelling buildings and construction plots. Other types of real estate are not reported mainly because of the required degree of homogeneity and sufficient number of observations in all different subgroups. Furthermore, apart from time and regions, the data also allows to examine possible dependence of prices on other factors such as the extent of wear or size of the municipality where the real estate is located, but these are not a concern of this thesis. Price developments in time are captured by the price index of Laspeyres type, which represent the ratio of two average prices for two different periods.

[2]For example, the publication from year 2016 includes data for years 2013, 2014 and 2015, with 2013 being the most precise.

Quarterly panel data for average purchase unit prices for flats were collected from a total of 16 publications, by combining quarterly price indexes with average prices in the base year.[3] The final dependent variable used in the analysis represents its deflated counterpart, i.e. real flat price, and will be denoted **rprice**.[4] Series for all region are available in Figure 3.1.

Figure 3.1: Real Flat Prices



3.2 Price Determinants

This section will present a list of factors that will be further used in empirical analysis. Apart from interest rate (collected from CNB), all series were collected from CZSO. As economic theory suggests, real estate prices should be related to both supply and demand factors. Previous works have employed various combinations of price determinants (e.g. Égert and Mihaljek, 2007; Hlaváček and Komárek, 2009; Belke and Keil, 2018) and the choice for this thesis was strongly limited by data availability. Firstly, factors will be described in more detail and expectations for relationship with housing prices will be set. Secondly, a summary of properties and descriptive statistics of all determinants will be given in Table 3.1. Thirdly, for better insight, graphs for all variables will be available in appendix in Figure 1 and Figure 2. It is also important to note that all nominal series were deflated to real ones with use of Consumer

[3]Indexes were recalculated to base 2017 and multiplied by 2017 average prices. Appropriateness of using average prices per squared meter was also confirmed by authors of publications, as they observed independence of prices on size of flats.

[4]in 2010 prices, CZK/m^2

Price Index (CPI), meaning they are in 2010 prices. Population adjustments are based on mid-year population in each region.

3.2.1 Labour Market Factors

rwage - Average real monthly wage. In quarterly analysis, time series had to be adjusted for seasonality.[5] Otherwise, results might be misleading, as the seasonality can mask the underlying trend.[6] Overall, it is considered to be a primary real estate price determinant, as it directly affects consumers disposable income and his savings, meaning higher wages should lead to higher demand for housing. Apart from that, it also indirectly affects the availability of mortgages.

unemployment - General unemployment rate. A negative relationship is expected as higher unemployment lowers disposable income and makes mortgages less available, which then leads to a decrease in demand for real estate.

activity - Economic activity rate of population. It represents the percentage of economically active persons in the population aged 15 and over. The sign is expected to be positive, interpretation is similar as in the case of lower unemployment.

3.2.2 Demographic Factors

natural, net migration, marriages, divorces - Natural increase in population, net migration, marriages and divorces respectively. All these factors affect housing demand via changes in the number of participants in the market and are therefore expected to have positive signs. Natural increase could be a little misleading since it reflects newborns, but it is expected to carry important information about the long term development in population. Net migration should be much less biased, as new inhabitants usually have no other option than to seek a new dwelling. Divorces and marriages should both lead to forming new households, as they either split or form a new one.

[5] This was done via X-13-ARIMA-SEATS and R package "seasonal".

[6] It has a simple explanation, since at the end of the year, wage rolls also include various bonuses.

3.2.3 Financial Market Factors

rinterest - Real yield rate on 10-year Czech government bonds. To get yield in real terms, inflation was deducted from nominal interest. It should reflect long term interest environment, and close link to mortgage financing costs would suggest a negative relationship with prices, as the higher costs reduce demand for housing.

rmortgage - Increase in real mortgage loans for housing. Higher volume of mortgages should indicate lower liquidity constraints, which tend to increase the availability of houses, leading to higher demand and a rise in prices.

3.2.4 Supply Factors

new flats - Number of completed flats. It should measure construction activity and reflect dynamics in the supply. Expectations are not straightforward. For example, a negative sign would represent an increase of supply relative to demand (prices decrease), while a positive relationship is also reasonable, as it could reflect the reaction to an increase in demand.

rconstrucion - Index of real construction work costs. Index is calculated from realized construction costs and includes expenditures on materials, wages, insurances and machinery. The index was also qualitatively adjusted and therefore does not reflect technological developments. It should capture the long term dynamics in flat construction costs, which are the main concern of developers. Therefore, a direct link to prices of new flats is expected, which affects changes in supply. Higher costs should then be positively related to flat prices.

Table 3.1: Data Summary

	Unit	Frequency	Panel	Mean	Std. Dev.
rprice	CZK/m^2	Q	✓	16052	8139
Labour market					
rwage	CZK	Q	✓	21103	3597
unemployment	%	Q	✓	6,57	2,88
activity	%	Y	✓	59,19	1,63
Demographic					
natural	‰	Y	✓	-0,18	1,17
net migration	‰	Y	✓	1,70	4,02
marriages	‰	Y	✓	4,80	0,48
divorces	‰	Y	✓	2,83	0,43
Financial market					
rinterest	%	Q	✗	1,33	1,69
rmortgage	mil CZK	Y	✗	102789	55077
Supply					
new flats	‰	Y	✓	2,72	1,13
rconstruction	index	Q	✗	97,57	4,06

Note: rprice, rwage, rinterest, rmortgage and rconstruction are nominal variables adjusted for inflation (in 2010 prices); Index has a base year of 2010; Q - quarterly, Y - yearly; Panel distinguishes between panel data and time series; ‰ denotes per 1000 inhabitants.

From Figure 1 and Figure 2 we can see, that Prague, Středočeský and Jihomoravský tend to be outliers in a lot of series (namely rprice, rwage, unemployment, net migration and new flats), and this fact should be taken into account in the analysis, especially during the interpretation of results. Also, there are many other variables used in literature, but unfortunately not available at the regional level of quarterly frequency in the Czech Republic. This may be the main shortcoming of empirical analysis conducted in this thesis, as some of these factors might play an important role in determining housing prices. Most commonly used are average rents, which allow comparing prices with future cash inflows (see for example Čadil, 2009). In this case, in spite of this statistic being monitored by IRI, the data is not publicly accessible. Apart from that, other variables that were not collected on the regional level, even though often considered important, are: GDP, income, housing stock, volume of mortgages and construction costs.[7]

[7] GDP and income were replaced by wages, aggregate data for mortgages and construction costs had to be used.

Chapter 4

Theory and Methodology

This section reviews methods that will be later applied in the empirical analysis. All econometric techniques aimed to utilize panel data, which should provide more information about the sample compared to traditional time series or cross-sectional data. First of all, before evaluating misalignments in prices on the real estate market, it was needed to conduct an analysis of fundamental price determinants. For this purpose, various approaches can be applied and a brief overview is given for example in Belke and Keil (2018). Two most commonly used are the present value approach and the model of supply and demand. The first comes from an idea of simple present value from finance, as it relates house prices to future rental cash inflows. The analysis is then done by evaluating the causal relationship between rents and prices together with stationarity of price to rent ratio (see for example Himmelberg et al., 2005; Zemčík, 2011). Alternatively, according to basic economic theory, real estate prices can be also modelled via the simple law of supply and demand. This approach was applied by for example Tsatsaronis and Zhu (2004); Rae et al. (2006); Égert and Mihaljek (2007); Hlaváček and Komárek (2009), and since the thesis also built on it, it will be explained in more detail.

Demand is typically considered a function of flat prices and other demand shifters, such as income, mortgage rate, demographic and market factors and expected rate of return (rents). Generally, housing demand D can be specified as following function:

$$D = f(\overset{+}{P}, DS) \tag{4.1}$$

where P stands for house prices, DS for vector of demand shifters and the sign indicates positive effect of change in prices on change in demand.

Supply S is then closely linked to profitability of house construction and can be specified as function of house prices and construction costs, including shifters such as land prices, worker wages and material costs:

$$S = f(\bar{P}, SS) \tag{4.2}$$

where SS stands for vector for supply shifters. If we then assume, that market is in equilibrium, prices clear supply and demand and P can be expressed as:

$$P = f(DS, SS) \tag{4.3}$$

This is why it is sometimes also called the approach of inverted demand, and it allows to evaluate changes in demand and supply simultaneously. Nevertheless, as Égert and Mihaljek (2007) emphasized, it does not, of course, imply that flat prices are always stable. Real estate market and especially its supply side, usually adapt to changes very slowly, which then leads to temporary deviation from long term equilibrium determined by fundamental factors. Therefore, it is also needed to distinguish between short run and long run dynamics. But most importantly, Equation 4.3 connects economic theory with models that will be further explained.

4.1 Fixed and Random Effects

This approach was mainly inspired by Hlaváček and Komárek and will be used as a benchmark for second model. Same slopes were assumed for all regressors across regions, and the model can be described by the following equation (Wooldridge):

$$y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{jit} + a_i + u_{it} \tag{4.4}$$

where $i = 1, ..., n$ represents i-th region, $t = 1, ..., T$ stands for time period and $j = 1, ..., k$ is then j-th explanatory variable. We usually refer to a_i as an unobserved effect or unobserved heterogeneity, that it is specific for each region and does not change over time, while u_{it} stands for idiosyncratic error that changes over time. Together $a_i + u_{it}$ make a composite error v_{it}, that is assumed to be $IID(0, \sigma_v^2)$.

According to estimation, if the individual effects do not exist ($a_i = 0$), then pooled OLS are efficient and consistent. Otherwise, Equation 4.4 has to be estimated in a different way, and in most cases, we conduct either fixed effects (FE) or random effects (RE). While FE examines the differences in intercepts and assume constant error variances across regions, RE assume the same intercepts and considers differences in error variances across regions. RE additionally assume that the regressors are uncorrelated with region-specific error ($Cov(x_{jit}, a_i) = 0, \forall j$). If this is satisfied, RE is more efficient than FE, otherwise RE is not consistent and inferior to FE. One might have better properties over the other, and we have to decide which one is more appropriate. This can be done either by economic intuition or via statistical testing.[1]

The verification of Classical linear model (CLM) assumptions is also needed. Presence of heteroskedasticity can be tested using the Breusch-Pagan test with a null hypothesis of homoskedasticity. It is actually not required for unbiasedness of OLS estimators, however, in presence of heteroskedasticity, standard errors and test statistics are no longer valid. Similarly, serial correlation also affects statistical inference and needs to be tested. This can be done with Breusch-Godfrey test with null hypothesis of no serial correlation in idiosyncratic errors. In the case of identification of heteroskedasticity and serial correlation, one has to estimate robust standard errors in order for our estimator to be BLUE. This can be easily done with Arellano (1987) method, which allows for a fully general heteroskedasticity and serial correlation structure.[2] The following chapter will explain the important concept of stationarity, that comes hand in hand with FE and RE.

4.2 Stationarity and Spurious Regression

In real world, macroeconomic time series are rarely stationary. However, regressing nonstationary variables generally gives spurious results, even if series are not trending. Under these conditions, stationarity of errors is uncertain, requirements for conducting OLS regression are not met, and there is a high risk that regression will indicate a non-existing relationship.[3] The concept of spurious regression was first described and simulated by Granger and Newbold

[1] Hausman test with the null hypothesis of consistent RE.
[2] in R this can be done via "coeftest" with Arellano covariance matrix
[3] The only exception is the concept of cointegration, which will be explained later

(1974). According to authors, these types of equations tend to have a high degree of fit (R^2), while also having a very low value of Durbin-Watson statistic.[4] In that case, the only conclusion about the results that can be made is, that the equation is misspecified. Therefore, before conducting the analysis, one has to test for unit roots in all included variables. The first step usually consists of plotting and examining the series. This allows to get an idea about their behaviour, form expectations about results and also provide some intuition about the correct setup.[5] Following formal tests for nonstationarity were employed and their selection was determined by the literature. Section explains only the key equations connected to each of them together with hypotheses, since the estimation process itself is usually complicated and not relevant for this thesis. The theory in the rest of this section was inspired mainly by EViews (2019) and Hurlin and Mignon (2007).

4.2.1 Time Series Tests

The first test that is commonly applied in literature is the Augmented Dickey-Fuller(ADF) (Dickey and Fuller, 1979). The test is built on an important assumption, that examined series follow autoregressive moving average $ARMA(p,q)$ process[6] and is based on estimating the following test regression:

$$\Delta y_t = \delta D_t + (\rho - 1)y_{t-1} + \sum_{j=1}^{\pi} \gamma_j \Delta y_{t-j} + \epsilon_t \tag{4.5}$$

where D_t are optional exogenous regressors, sometimes called deterministic components, and can be either constant or constant with trend. Estimated coefficients are δ and ρ, y_t is the examined series and tested hypotheses are:

$$H_0 : \rho = 1; H_1 : \rho < 1$$

The null hypothesis is that the process contains a unit root, i.e. is nonstationary, while the alternative suggests stationarity of the series. The lagged difference terms in sum in Equation 4.5 are supposed to approximate the $ARMA$ structure of the errors, to make them serially uncorrelated. However, correct selection of the lag length (π) can be cumbersome as the structure of the time series processes is always unknown.

[4] Signalling strong autocorrelated errors.

[5] Choice of deterministic components, number of lags, etc.

[6] p autoregressive terms, q moving average terms

Phillips and Perron (1988) introduced an approach that is closely linked to ADF. The difference is that it is nonparametric and robust to any general serial correlation and heteroskedasticity in errors, thanks to Newey and West (1986) heteroskedasticity and autocorrelation consistent covariance matrix estimator. The hypotheses are the same as in case of ADF and the test regression is the following:

$$\Delta y_t = \delta D_t + (\rho - 1)y_{t-1} + \epsilon_t \tag{4.6}$$

Apart from robustness, another advantage of the Phillips-Perron (PP) test is that there is no need to specify the lag length for the test regression. Nevertheless, in spite of both ADF and PP tests being asymptotically equivalent, there might be significant differences in finite samples, because of the differences in correcting for serial correlation (see for example Schwert (1989)).

4.2.2 Panel Data Tests

Thanks to recent developments in panel analysis, series can be analysed in more detail. Therefore, various panel unit root tests, which generally have higher power than traditional individual time series tests, will be employed. They are very similar to time series tests, but applied on panel structures and sometimes need additional assumptions. Firstly, section introduces tests with a common unit root (LLC, Hadri), secondly with individual unit roots (IPS) and lastly addresses the problem of cross-sectional dependence via CIPS. Each test has both advantages and disadvantages, assessing their results simultaneously should give more insight into the true structure of the series.

Levin et al. (2002) developed a procedure in which t-statistics are pooled and is considered very similar to ADF equation as in time series (Equation 4.5). The only difference is the index i, which allows for pooling of our observations:

$$\Delta y_{it} = \delta D_{it} + (\rho - 1)y_{i,t-1} + \sum_{j=1}^{\pi} \gamma_{ij} \Delta y_{i,t-j} + \epsilon_{it} \tag{4.7}$$

And again, the hypotheses are:

$$H_0 : \rho = 1; H_1 : \rho < 1$$

In other words, under the null hypothesis the unit root is present in all individual series against the alternative that each is stationary. Homogeneous autoregressive coefficients (ρ) are assumed to be the same,[7] whereas lag lengths (π) and deterministic components (D_{it}) are allowed to vary among regions. Apart from assumptions, another weakness is that similarly to ADF, lags need to be specified apriori.

Hadri (2000) introduced another type of residual-based test for heterogeneous panels and considered the following model:

$$y_{it} = r_{it} + \beta_i t + \epsilon_{it} \tag{4.8}$$

where $r_{it} = r_{i,t-1} + u_{it}$, i.e. r_{it} is random walk and t is the optional trend component. Since ϵ_{it} are assumed to be IID, null hypothesis of stationarity is then $\sigma_u^2 = 0$. To be more precise, the null hypothesis states that the time series for each cross-section unit is stationary, while the alternative is that at least one time series has a unit root. The main advantage over LLC is that it allows for heteroskedasticity in disturbance terms across regions. Deterministic components are also allowed to vary, but there is not an option of no exogenous regressors.

Im et al. (2003) considered the same test regression as Equation 4.7 with an important difference in unit root term as it is now assumed to vary among regions:[8]

$$\Delta y_{it} = \delta D_{it} + (\rho_i - 1)y_{i,t-1} + \sum_{j=1}^{\pi} \gamma_{ij} \Delta y_{i,t-j} + \epsilon_{it} \tag{4.9}$$

$$H_0 : \rho_i = 1, \forall i$$

$$H_1 : \rho_i < 1, \exists i$$

In other words, rejection of null hypothesis would mean that the series in some regions have no unit root. Similarly to LLC, we also need to specify the number of lags. Deterministic components are different for each region and have similar limitations as in the case of Hadri.

[7] Assumption of a common unit root.
[8] Assumption of individual unit roots.

All panel tests introduced above share an important assumption of cross-sectional independence. However, in most cases, this is very unlikely to be satisfied, especially in the case of regional analysis, as regions are strongly interconnected and dynamics are very likely to be correlated.[9] If the independence is wrongly assumed, panel unit root tests can lead to spurious results. Therefore, one additional test will be introduced to address this issue.[10]

Pesaran (2007) proposed a cross-sectionally augmented version of the already described IPS test, that accounts simultaneously for cross-sectional dependence and autocorrelation. It extends IPS presented earlier by adding the cross-section average of lagged levels together with averages of lagged first differences to standard ADF equation, and has the following structure:

$$\Delta y_{it} = \delta D_{it} + (\rho_i - 1) y_{i,t-1} + \theta_i \bar{y}_{t-1} + \sum_{j=0}^{\pi} \phi_{ij} \Delta \bar{y}_{i,t-j} + \sum_{j=1}^{\pi} \gamma_{ij} \Delta y_{i,t-j} + \epsilon_{it} \quad (4.10)$$

After obtaining cross-sectionally augmented Dickey-Fuller (CADF) statistics ($t_i(N,T)$) given by t ratio of the coefficient ρ_i for each region, the CIPS test statistics for the whole panel is then calculated in the following way:

$$CIPS(N,T) = \frac{1}{N} \sum_{i=1}^{N} t_i(N,T)$$

Hypotheses are the same as in the case of IPS, and again, CIPS also suffer significantly from a wrong selection of lags.

By applying these tests on levels and differences of variables, one can determine the order of integration of the series. The following cases can arise:

1. Stationary in levels - Integrated of order zero (I(0))
2. Nonstationary in levels
   (a) Stationary in first differences - Integrated of order one (I(1))
   (b) Nonstationary in first differences - A higher level of integration

If nonstationary series are found, namely I(1) processes, the usual treatment is then to take first differences or growth rates of all nonstationary variables

[9] This can be seen in the data description: Figure 1, Figure 2.
[10] We usually refer to group of these tests as a second-generation unit root tests.

or include lag of our dependent variable. On one hand, spurious regression does no longer occur, but on the other hand, important information about the long-run relationships is lost, which is the main limitation of this approach. Additionally, the dimensions of our datasets are relatively short and therefore we need to be careful with the interpretation of results.[11] Thesis aims to address these issues using an alternative approach, which will be described in the rest of this chapter.

4.3 Cointegration

As was already mentioned, there exists one special case in which regression of nonstationary variables is not spurious. That is when variables are in a cointegrating relationship. It basically means that if nonstationary variables are truly related, which could be suggested by economic theory, they will "move together" as their stochastic trends should be very similar. If these series have the same order of integration, there will exist their linear combination which is in fact stationary.

According to empirical tests for cointegration, the so-called "two-step" procedure introduced by Engle and Granger (1987) can be conducted. Their method is applied after testing for stationarity in variables and is conditional on the same order of integration in all variables.[12] In the first stage of Engle and Granger method, an I(1) dependent variable is regressed on all I(1) independent variables and the estimated residuals are saved. In the next stage, we then test saved residuals for unit roots. If these residuals are found to be stationary, variables tend to return to equilibrium after shock and the long run relationship is confirmed. On the other hand, if our series are not cointegrated without any long run relationship, then the residuals should have a unit root. Engle and Granger procedure was applied on panel data by Pedroni (2004), Kao (1999) and Westerlund (2007) and their approaches will be now briefly introduced.

[11] In all panel unit root tests, asymptotic behaviour of at least one of the dimensions has to be assumed.

[12] I(1) in my case

Pedroni presented a test in which he assumes heterogeneous intercepts, slopes and trend coefficients and its test statistics are based on the following equation:

$$y_{it} = \alpha_i + \delta_i t + \sum_{j=1}^{k} \beta_{ji} x_{jit} + \epsilon_{it} \tag{4.11}$$

where $\alpha_i + \delta_i t$ are the optional deterministic components and k is the number of included explanatory variables. Residuals $\hat{\epsilon}_{it}$ are saved and tested for unit roots:[13]

$$\hat{\epsilon}_{it} = \rho_i \hat{\epsilon}_{it-1} + v_{it} \tag{4.12}$$

$$H_0 : \rho_i = 1, \forall i$$

$$H_1 : \begin{cases} \rho_i = \rho < 1, \forall i \text{ (within)} \\ \text{or} \\ \rho_i < 1, \forall i \text{ (between)} \end{cases}$$

Rejection of the null hypothesis of no cointegrating relationship would mean accepting one of two alternative hypotheses:

1. Series are cointegrated with homogenous ρ (within-dimension test)
2. Series are cointegrated with heterogeneous ρ (between-dimension test)

In the first setup, panel statistics from 3 non-parametric tests[14] and 1 parametric test[15] are provided. In the case of the second less restrictive alternative hypothesis, the reported group statistics are the same except they exclude the variance ratio.

Kao procedure for testing cointegrating relationships is very similar to Pedroni and again, all variables are assumed to be nonstationary. Kao is based on modification of Equation 4.11, which assumes slope coefficients to be homogenous among cross-sections and trend coefficients to be equal to zero:

$$y_{it} = \alpha_i + \sum_{j=1}^{k} \beta_j x_{jit} + \epsilon_{it} \tag{4.13}$$

Also ρ_i from the auxiliary regression 4.12 is assumed to be the same for all cross-sections ($\rho_i = \rho$).[16] The hypotheses are a little different as now the re-

[13] As in the case of ADF, this equation can also be augmented with lags.
[14] v-statistic, Phillips-Perron ρ-statistic, Phillips-Perron t-statistic
[15] ADF-statistic
[16] Again, it can be augmented with lags of residuals.

jection of null means that all cross-sections are cointegrated, and only one test statistic is reported.[17]

So far, all introduced cointegration tests assumed cross-sectional independence, but as in the case of testing for unit roots in the previous section, it might be unrealistic for my panel. Therefore, this issue will be also addressed by second-generation cointegration test.

Westerlund introduced error correction based test, in which the problem of cross-sectional dependence is addressed via robust P-values generated by bootstrapping. Also, in comparison to residual-based test (Pedroni or Kao), the restriction of common factor is relaxed by allowing long term and short term parameters to be heterogeneous. Equation itself will not be explained, but is very similar to one that will be introduced in the following section and as in case of Pedroni, either no deterministic component, constant or constant with trend can be included. The cointegration hypotheses are then formed based on the error correction terms (ECT) calculated for each cross-section. Not rejecting the null hypothesis of ECT being equal to zero for all cross-sections would imply no cointegrating relationship. Regarding the alternative hypothesis, 2 types are considered:

1. Group alternative - At least one cross-section is cointegrated

2. Panel alternative - The whole panel is cointegrated

For testing purposes, 4 tests were developed by Westerlund, i.e. 2 for each type of alternative hypothesis.[18]

4.4 Error Correction Model

Until now, especially because of the short length of series, problems such as endogeneity were not addressed. Also, I was able to only estimate short run relationship as the results would be otherwise spurious. Therefore, I will also proceed to a more complex model, which will be estimated from data of higher frequency and is expected to have greater statistical power than the model introduced in Section 4.1.[19] The methodology used is similar to Égert and Mi-

[17] ADF-statistic

[18] Least squares estimate of ECT and t-ratio of ECT.

[19] The reason why a similar procedure was not applied to annual data set as well is that cointegration tests are very likely to give unreliable results for short lengths of series.

haljek (2007), Bill and Ivarsson (2009) or Stohldreier (2012). It is important to note that the appropriateness of this model is conditional on the cointegrating relationship. If tests from the previous section would indicate no cointegration, the steps that will be described further could not be followed. The first part will focus on the estimation of the long run coefficients of the cointegrating (equilibrium) relationship and finding of the fundamental flat price determinants. The second part will then connect it to short run dynamics and evaluate the speed of adjustment to equilibrium via ECT.

4.4.1 Long Run

So far, I was only able to determine whether the variables are cointegrated or not and I have not yet examined the actual long run effects of explanatory variables. Such long run steady state relationship can be described by the following equation:

$$y_{it} = a_i + \sum_{j=1}^{k} \beta_j x_{jit} + u_{it} \tag{4.14}$$

where a_i are regional fixed effects and x_{jit} are both supply and demand determinants of flat prices suggested by the literature. It is essentially the same as Equation 4.4, however, OLS cannot be applied, as the estimates would be biased and would only lead to false inferences. The reason is that in the presence of cointegrating relationships, error terms are very likely to be correlated with regressors, since all variables share a common stochastic trend. Therefore, the endogeneity problem need to be addressed, which can be done in different ways. In literature, panel dynamic OLS (DOLS) or panel fully modified OLS (FMOLS) estimators usually prevail. Panel DOLS address it by adding leads and lags of the first differences of the explanatory variables to Equation 4.14 and then estimating it by standard OLS. Whereas panel FMOLS reduce this bias with a nonparametric approach, which is, in fact, more computationally demanding and harder to understand. Nevertheless, both methods will not be utilized as they were proven to be asymptotically the same and it would only complicate the final interpretation. Estimates will be obtained solely by panel DOLS, since generally, they offer better properties in finite samples. Also, heteroscedasticity and autocorrelation consistent (HAC) standard errors[20] will be considered to improve the robustness of results.

[20]Newey-West standard errors

Finally, the model from this section will not only be used for examining the effects of individual determinants, but also for the residual analysis, which will aim to evaluate the possible price bubbles.

4.4.2 Short Run

Here, focus is given to the short run process of correcting for deviations from the equilibrium that was estimated previously. If there actually exists some long run relationship, there also needs to be some mechanism, that returns short run misalignments back to their equilibrium. To connect short run with long run, Error Correction Model (ECM) will be conducted.[21] To be more precise, this short run adjustment mechanism will be captured by estimating the following equation:

$$\Delta y_{it} = b_i + \sum_{j=1}^{k} \gamma_j \Delta x_{jit} + \Phi ECT_{i,t-1} + v_{it} \tag{4.15}$$

where b_i is the region-specific fixed effect and the Error correction term ($ECT_{i,t-1}$) is the lagged residual from Equation 4.14 ($\hat{u}_{i,t-1}$), in other words, this holds:

$$ECT_{i,t-1} = y_{i,t-1} - \hat{a}_i - \sum_{j=1}^{k} \hat{\beta}_j x_{ji,t-1}$$

The reason behind adding ECT was to evaluate the speed of adjustment to equilibrium via its coefficient Φ. Its value is expected to be between -1 and 0, using the intuitive understanding that the closer to -1, the faster is the adjustment process. To give some intuition, if determinants did not change and houses were overpriced in previous period, in order to return to equilibrium, we would need prices to decrease, which would imply $\Phi < 0$. The lower boundary of -1 is then given by the fact that lower values would mean that the relationship does not converge, but rather "jumps" from positive to negative disequilibrium and vice versa. Nevertheless, there is no limitation for its estimates to be in expected range. Values lower than -1 or higher than 0 would imply no convergence to equilibrium in the long run and indicate some serious instabilities.[22]

[21] Long term relationship measures the effects of variables in levels while short run is represented by their differences.

[22] Some authors also consider values $(-2; -1)$ sufficient for oscillatory convergence (see for example Loayza and Ranciere (2004))

It is also important to note that all variables in Equation 4.15 are in fact stationary[23] and therefore, problems of spurious regression were avoided. The estimation itself will be done by FE or RE[24] as introduced in Section 4.1 and again, the presence of autocorrelation and heteroskedasticity will be tested and eventually corrected.

[23]Stationarity of ECT was confirmed by cointegration tests.
[24]Based on Hausman test

Chapter 5

Results

5.1 Fixed and Random Effects Analysis

The first concern was to determine the order of integration of all variables to avoid spurious regression. Therefore, the unit root tests described earlier were applied, and summarized the results in Table 1 and Table 2. The reason why some variables were tested by time series tests is that they are the same for all regions and conducting panel tests would not make any sense. Based on corresponding p-values for different types of hypotheses, the order of integration for each variable was determined, and findings were included in the last column of both tables. Generally, the decision was straightforward, but for some variables (especially the demographic factors), few tests were in contradiction, and in such cases, economic intuition was also considered.[1] Such outcome is the disadvantage of conducting multiple tests at the same time, however, the results are surely more robust to possible misspecification. Overall, it can be concluded, that all variables are stationary in their differences. Results from unit root tests on their levels also seem to be very reasonable, as they are in line with literature.[2]

As suggested earlier, the next step was to take first differences of all I(1) explanatory variables and only then include them in regression. Unit root in dependent variables was addressed by two alternative approaches as suggested, for example, by Hlaváček and Komárek. Firstly, the I(1) dependent variable was kept, but additionally its one year lag was included as another explanatory

[1] In the case of demographic variables, they already represent a differenced value.

[2] The only difference is the order of integration of unemployment, which clearly appears to be nonstationary in my analysis, which contradicts for example Hlaváček and Komárek.

variable (Equation 5.1) and secondly, all nonstationary variables (including flat prices) were differentiated (Equation 5.2).

$$\begin{aligned} rprice_{it} =& \beta_0 + \beta_1 \Delta rwage_{it} + \beta_2 \Delta unemployment_{it} + \beta_3 \Delta rmortgage_{it} \\ & + \beta_4 \Delta rconstruction_{it} + \beta_5 new\_flats_{it} + \beta_6 natural_{it} \\ & + \beta_7 net\_migration_{it} + \beta_8 marriages_{it} + \beta_9 divorces_{it} \\ & + \beta_{10} activity_{it} + \beta_{11} rinterest_{it} + \beta_{12} rprice_{i,t-1} + a_i + u_{it} \end{aligned} \tag{5.1}$$

$$\begin{aligned} \Delta rprice_{it} =& \beta_0 + \beta_1 \Delta rwage_{it} + \beta_2 \Delta unemployment_{it} + \beta_3 \Delta rmortgage_{it} \\ & + \beta_4 \Delta rconstruction_{it} + \beta_5 new\_flats_{it} + \beta_6 natural_{it} \\ & + \beta_7 net\_migration_{it} + \beta_8 marriages_{it} + \beta_9 divorces_{it} \\ & + \beta_{10} activity_{it} + \beta_{11} rinterest_{it} + a_i + u_{it} \end{aligned} \tag{5.2}$$

The procedure described in Section 4.1 was then applied to both models and estimation itself was done by FE,[3] since the Hausman test rejected consistency of RE in both equations (see Table 3). Also, the presence of heteroskedasticity and serial correlation was indicated by Breusch-Pagan and Breusch-Godfrey tests respectively, and therefore robust standard errors were calculated instead. The final estimation results are available in Table 5.1, standard errors were not included for space-saving purposes.

It is not surprising that R^2 from the first model is much higher than from the second one, since lagged prices clearly explain a lot of variation. Nevertheless, R^2 of 0.418 is still substantial for macroeconomic regression and the overall fit si considered to be very decent. Generally, both models appear to be very similar despite the different method of taking care of nonstationarity in the dependent variable. Sign wise, the only disparities were found in natural and divorces variables, but since the first mentioned is statistically significant on a very low level, no logical explanation could actually cause doubts about the correctness of the second model. Significance wise, Δrmortgage, marriages and divorces showed some dissimilarities, but since all of them were found to be significant only on 10% level, this is not much of a concern.

[3]Fixed effects have a different interpretation. While in Model 1, they represent differences in price levels across regions, in the second they reflect differences in price dynamics.

Table 5.1: Yearly analysis: Results

	Model (1)	Model (2)	
	Dependent variable:		
	rprice	Δ rprice	Expected sign
rprice $_{t-1}$	0.684^{***}		+
Δ rwage	0.668^{***}	0.857^{***}	+
Δ unemployment	-245.802^{**}	-337.632^{***}	−
Δ rmortgage	0.001	0.009^{*}	+
Δ rconstruction	−24.992	−48.577	+
new flats	−124.875	−238.701	+/−
natural	276.342^{**}	-249.259^{***}	+
net migration	295.212^{***}	248.931^{***}	+
marriages	268.678	737.071^{*}	+
divorces	-521.815^{*}	160.570	+
activity	105.247	59.073	+
rinterest	−63.945	−66.000	−
Observations	238	238	
R^2	0.820	0.477	
Adjusted R^2	0.799	0.418	

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ *. Estimated by FE. Arellano robust errors. Author's calculations.*

Most importantly, I will comment on significance levels of other variables that behave similarly in both models. Models clearly suggest that the fundamental determinants of real flat price developments are Δrwage, Δunemployment, and two demographic factors: natural and net migration. Again, this is in line with the literature (for example Égert and Mihaljek, Hlaváček and Komárek, Belke and Keil), but additionally, strong connection with net migration, which appears to explain a lot of price variability, is observed.[4] Significant values for lagged rprices were expected, as flat prices are assumed to be strongly persistent. Very surprising results were given by the fact that both supply determinants (new flats, rconstruction) turned out not to be significant. Naturally, this could show the unimportance of supply side in terms of determining flat prices, but we also have to keep in mind that only 2 factors were included and both play role rather of proxies, as it was very challenging to measure supply with available data.

[4] Especially in the regions of Prague or Středočeský (see Figure 1).

The individual magnitudes of regression coefficients will not be interpreted, as they are not very interesting, but overall, all of them can be considered economically significant.[5] [6] However, signs of these coefficients are much more interesting. For more clarity, expected signs were included in the fourth column of Table 5.1, and the intuition behind them was already given in Section 3.2. The vast majority of signs of estimated coefficients actually match expectations, which confirms the appropriateness of both models for further use in the residual analysis. The only discrepancies appeared in the case of insignificant Δ rconstruction or in case of natural (Model 2) and divorces (Model 1).

In terms of regional differences, there is not much that can be evaluated right now, since in both models, homogeneous slope coefficients among regions are assumed. However, from F-test can be concluded that fixed effects in both models are significant, which indeed confirms some degree of heterogeneous behaviour. A more detailed analysis of residuals with respect to regions will be performed later.

5.2 Error Correction Model

Before estimating both long run and short run, unit roots[7] as well as cointegration had to be tested. Due to limited data availability, not as many variables as in yearly analysis were included. The same stationarity tests as in Section 5.1 were performed and results are available in Table 4 and Table 5. Except for rinterest, all tests suggest the same conclusions about the presence of unit roots in series as in Section 5.1. This is actually a very interesting result, which partially confirms the viability of the analysis conducted on the data of lower frequency. All variables were found to be I(1), but instead of the first differencing, analysis proceeded to test for cointegrating relationships between rprice, rwage, unemployment, rconstruction and rinterest. The statistics and p-values from all 3 tests described earlier are available in Table 5.2.

[5] Keep in mind that they represent an effect on prices per m^2, the standard flat is assumed to have $68m^2$

[6] Interpretation is a little complicated as some variables were regressed in their first differences and not in levels. For example: increase of 1000 CZK in Δ rwage (in other words rwage accelerates by 1000 CZK) is expected to increase rprices by 668 CZK/m2 (model 1) or increase Δ rprice (or accelerate rprice) by 857 CZK/m2 (model 2) , ceteris paribus).

[7] Cointegration requires all variables to be I(1)

Table 5.2: Cointegration results

Pedroni			Kao			Westerlund		
Panel						*Group*		
v-stat	3,650	0,000	t-stat	-3,630	0,000	Gt	-3,666	0,003
rho-stat	0,784	0,783				Ga	-1,049	0,005
PP-stat	0,498	0,691				*Panel*		
ADF-stat	-4,548	0,000				Pt	-3,759	0,010
Group						Pa	-2,089	0,003
rho-stat	2,530	0,994						
PP-stat	1,975	0,976						
ADF-stat	-6,689	0,000						

Note: Statistic on left, p-value on right. Deterministic component - intercept. Lag length - AIC. Westerlund robust P-values - 400 bootstrap replications. Coloured values indicate cointegration at 5% significance level. Author's calculations.

Various setups were tested and the results were found to be very sensitive to the choice of parameters. Since I have no rationale especially for the selection of leads and lags, I opted for determining their length based on information criterion[8], and selection of deterministic components was based on graphs/intuition. This might be the main reason why the results are quite ambiguous, however I still believe that a meaningful conclusions can be made. Even though Pedroni suggested no cointegrating relationship with heterogeneous AR coeficients[9] and is unclear in case of cointegration with homogeneous AR coefficient,[10] results from Kao and Westerlund are much more promising. Both confirmed strong cointegrating relationship, while the latter one is considered the most reliable out of all. The underlying reason is that it is robust to cross-sectional correlation, and since all four alternative hypotheses were accepted (especially the panel statistics), it was concluded, that the relationship is indeed cointegrated, and it is appropriate to conduct the long run model, which can be described by the following equation:

$$\begin{aligned} rprice_{it} = & a_i + \beta_1 rwage_{it} + \beta_2 unemployment_{it} \\ & + \beta_3 rconstruction_{it} + \beta_4 rinterest_{it} + u_{it} \end{aligned} \tag{5.3}$$

Results from panel DOLS estimation are summarized in Table 5.3. Standard

[8] Akaike information criterion (AIC) or Schwarz information criterion (SIC), max lags based on length of series

[9] Accepts only 1 out of 3 group alternative hypotheses.

[10] Accepts 2 out of 4 panel alternative hypotheses.

errors and corresponding p-values robust to autocorrelation and heteroskedasticity were additionally included.

Table 5.3: Quarterly analysis: Long run results

	Dependent variable: rprice		
	DOLS	Robust DOLS	Expected sign
rwage	1,078***	1,078***	+
	(0,061)	(0,091)	
unemployment	−224,188***	−224,188*	−
	(82,586)	(121,774)	
rconstruction	323,002***	323,002***	+
	(29,883)	(44,063)	
rinterest	−294,685***	−294,685*	−
	(103,948)	(153,273)	
Observations	970	970	
R^2	0.133	0.133	
Adjusted R^2	0.116	0.116	

*Note: *p<0.1; **p<0.05; ***p<0.01 . Estimated by panel DOLS (EViews). Newey-West ("hac") errors in case of Robust DOLS. Individual selection of leads and lags done by AIC (max 1 lag and 1 lead). Deterministic component - constant. rwage seasonally adjusted. Author's calculations.*

A very high value of adjusted R^2 could actually raise some concerns about the viability of such specification. However, very similar results were found by for example Égert and Mihaljek and should not be very surprising in the case of modelling cointegrating relationship. All variables turned out to be statistically significant, which suggest that all of them are fundamental determinants of flat prices. In case of robust errors, the significance of unemployment and rinterest slightly dropped, but they are still significant on 10% level. All signs of coefficients match their expected counterparts and it can be concluded that these have base in economic theory.

Compared to yearly analysis, the main differences lie in significance levels. Now, rinterest and rconstruction also turned out to be fundamental flat price determinants. The reason why such relationship was not detected is very likely to be given by the lower frequency of data. Estimated coefficients have also much better properties as they now describe long run relationship of variables in levels and did not lose potentially important information from first differencing. Additionally, all of them have a more straightforward interpretation and appear to be economically significant.

Next, the following short run equation, in which the ECT obtained from long run relationship, was estimated:

$$\begin{aligned}\Delta rprice_{it} =& b_i + \gamma_1 \Delta rwage_{it} + \gamma_2 \Delta unemployment_{it} \\ &+ \gamma_3 \Delta rconstruction_{it} + \gamma_4 \Delta rinterest_{it} + \Phi ECT_{i,t-1} + v_{it}\end{aligned} \tag{5.4}$$

The estimation was performed by FE as Hausman rejected consistency of RE (Table 3) and autocorrelation and heteroskedasticity was addressed via robust errors (similarly as in yearly analysis). Results are published in table Table 6. The main interest was the estimation of ECT coefficient, which turned out to fulfill both criteria needed for convergence, as it is statistically significant and lies in the desirable range. Its magnitude of $-0,043$ suggests a very slow adjustment process, in other words, approximately $4,3\%$ of deviation from equilibrium is corrected in each following quarter. In comparison, for instance findings by Égert and Mihaljek suggest the ECT to be between -0,099 and -0,046 in case of OECD or between -0,361 and -0,146 in case of CEE regions.[11] All other coefficients are also significant and do have expected signs. Again, the focus is not given to their interpretation, but in general, the differences in magnitudes in comparison to models from Section 5.1 are not negligible.

Robustness of results from both long and short run models was also increased by conducting various specifications, that did not produce significantly different results.[12] The quarterly analysis is expected to be more accurate, nevertheless, the yearly analysis appeared to produce reliable results that should be used along with ECM for the examination of price bubbles in the following section.

5.3 Housing Price Bubbles

5.3.1 Simple Economic Indicators

This section aims to build expectations about what periods should be identified as a bubble, based on price to income (P/I) and price to rent (P/R) indicators. Even though CNB monitors their more advanced counterparts, I will draw from

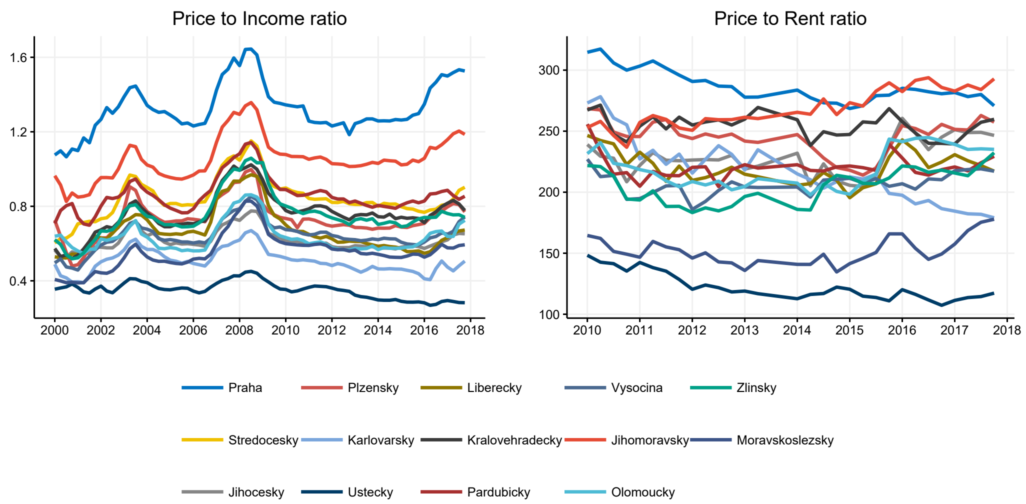
[11] Ranges are given by different variable specifications.

[12] For example the exclusion of Prague and Jihočeský, or some determinants such as rinterest or rconstruction. Results not published for space-saving purposes.

my own calculations and employ the available dataset.[13] Higher values of P/I are connected with lower housing affordability, while higher R/I values indicate that rental should be preferred over owning. In other words, higher values of both of these indicators show a higher likelihood of flat price overvaluation.

Firstly, the average levels of both indicators with respect to regions will be described. There are definitely some outliers, namely Prague and Jihomoravský with high values, while region Ústecký experienced the lowest ones in cases of both ratios. Secondly, "sudden increases" in these ratios should suggest bubble periods. As we can see, P/I identified 2 bubbles in all regions in 2003 and 2008, as well as a significant increase in 2017 in Prague and Jihomoravský. According to R/E, not much can be said, as data is available only for a very limited period, and the ratio seem to be stable over time. In spite of these indicators being widely used and certainly giving some intuitive understanding, they fail to take into account the dynamics of other fundamental determinants, which will be addressed in the following section.

Figure 5.1: Bubble Indicators



Note: Data for rents available only from 2010 with missing values for second half of 2013. Calculations are based on data from IRI and were performed in CNB. Graphs generated in R.

[13]Calculations are done by simply dividing these variables

5.3.2 Price bubble analysis

Finally, models from Section 5.1 and 5.2 will be utilized for evaluation of price bubbles in the whole period from 2000 to 2017. For this purpose, yearly Model 1 with lagged dependent and quarterly model of a long run relationship were selected. Since they both share the same dependent variable and each potentially offer additional information, they will be evaluated simultaneously.

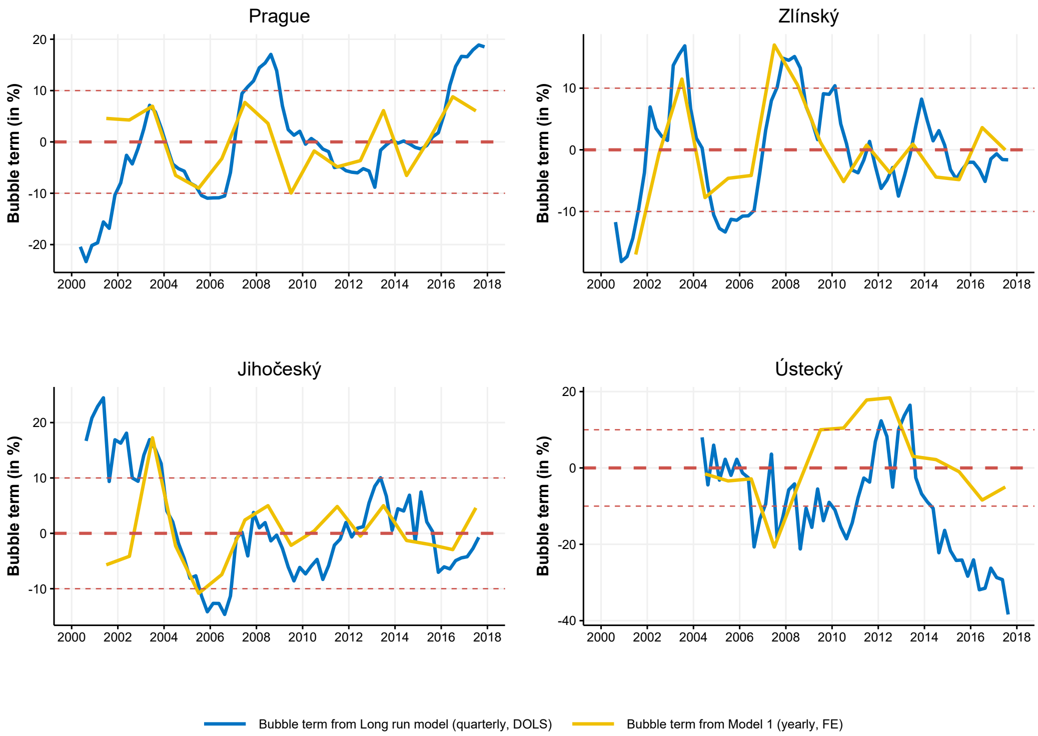
The examination itself can be done in many ways, in most cases we examine residuals, i.e. the difference between the actual observed prices and the estimated prices, and analyse their behaviour. But for my panel, in which price levels differ significantly among regions, such approach is likely to be misleading. Some authors also depict fitted values and compare them with observed prices and even though this approach somehow address the issue discussed earlier, it gets very confusing when comparing two models at the same time. Therefore, I will make use of the so-called "bubble term" (Equation 5.5), which is defined as a percentage difference between the actual and estimated prices. From an economic point of view, it represents percentage deviation from equilibrium.

$$B_{it} = \frac{rprice_{it} - \widehat{rprice}_{it}}{\widehat{rprice}_{it}} * 100\% \tag{5.5}$$

where B_{it} is the bubble size for region i in time t, $rprice_{it}$ is the actual flat rprice and $\widehat{rprice}_{it}$ is the fundamental (estimated) flat rprice. Based on this price bubble measurement, it was decided whether the bubble is significant or not. The decision was usually established on an artificial value of 10% which was supposed to make the results more understandable and has no deeper meaning.

Even though each region exhibits unique dynamics of the bubble term, I opted to split them into groups based on the underlying pattern of overvaluation periods. It turned out that 4 main groups emerged and for better insight, their representatives are depicted in Figure 5.2. Graphs for other regions are available in the appendix in Figure 3 and Figure 4. It is important to note that lengths of bubble periods vary by region. They usually range from 1 to 3 years, which was not considered during the sorting process, as it would create too many subgroups.

Figure 5.2: Real Estate Bubbles: Results



Note: Values of Bubble term were evenly distributed on the horizontal year axis for more clarity and better comparison, i.e. for quarterly data we set 2010 Q1 = 2010,125; 2010 Q2 = 2010,375; 2010 Q3 = 2010,625; 2010 Q4 = 2010,875, and for yearly: 2010 Y = 2010,5. Graphs generated in R.

Generally, both models struggle to produce the same values for the early as well as most recent years. Also, it appears that Model 1 (yellow line) tends to underestimate the deviations, because the values are most of the time below the Long run model (blue line). This can be seen especially in the regions of Prague, Středočeský and Plzeňský, where Model 1 underestimates the bubble in 2008. On the other hand, they usually both confirm the presence of a bubble in periods which were already suspected based on the P/I indicator. In all unclear cases, the outcome from Long run model was strictly preferred, as it appears to best explain the dynamics in prices.

The first group consists of regions of Prague and Jihomoravský. According to the analysis, these regions experienced the bubble during the crisis of 2008 as well as in 2017, however, Jihomoravský barely hit the boundary of 10%. They are the only ones that appear to face a high risk of current price overvaluation, as the bubble term's trend is expected to keep growing.[14] In the

[14]Prices increased since 2017

second group, there are regions of Zlínský, Středočeský, Plzeňský, Pardubický, Liberecký, Olomoucký, Moravskoslezský and Královehradecký. The results suggest that all experienced bubbles in 2003[15] and in 2008, while currently being almost in equilibrium. This majority group appears to be the best representative of aggregated values for the Czech Republic. The third group includes regions of Jihočeský, Karlovarský, Vysočina, which seem to experience bubbles only in the early stage of the monitored period. However, this need to be taken with caution, as model 1 indicates the opposite scenario of undervaluation in all 3 cases, and it is therefore possible that these regions did not experience any significant bubble over the whole period. Lastly, there is region Ústecký, which turned out to have very unique behaviour. Values of the bubble term for the period of 2000-2003 had to be removed from Figure 5.2, as they turned out to be extremely high. The reason is that Long run model estimated very low fitted values mainly because of the very high level of unemployment. The residuals were actually not much larger compared to other regions, but together with the lowest flat prices, it suggested an overvaluation of up to 3000% (in comparison to 10% suggested by Model 1). Such results were not found realistic, and it is questionable how reliable the values for the remaining years are, suggesting overvaluation in period of 2012-2013.

According to undervaluation, which is much less of a concern as it does not pose much danger for the economy, there are also some patterns. All regions appear to be undervalued during the pre-crisis period of 2006-2007, and Liberecký, Karlovarský and Ústecký also appear to be undervalued in recent years. Additionally, regions of Prague, Středočeský, Plzeňský, Královehradecký, Jihomoravský, Zlínský, seem to experience undervaluation in very early years.

These findings, however, cannot be directly compared with the literature, since to my knowledge, there has not been any similar regional analysis in recent years. Nevertheless, results from Hlaváček and Komárek, who analysed the period of 1998-2008, can be utilized. The models he employed together with simple indicators suggested housing price bubbles in 2003 and 2008, which partly confirms my results, even though bubble periods for individual regions generally differ. These discrepancies are likely to be caused by the extremely short length of his data set together with a higher probability of overfitting.

[15] Some experienced it already in 2000.

According to more recent studies or papers, use of CNB's Financial Stability Report 2018[16], can be made. CNB model indicated flat prices to be overvalued by approximately 30% in 2008 and 14% in 2017, both in the Czech Republic as well as in Prague, however values for other regions are not available. On the contrary, my results suggested a much lower degree of overvaluation in 2008 and similar values in 2017 only for Prague and Jihomoravský. It is unclear whether aggregating the data would lead to similar results, as this group is not expected to be the best representative of the Czech Republic. Another possibility are the Thematic articles on financial stability, also published by CNB. Plašil and Andrle (2019) applied two diametrically different approaches of borrowing capacity of households and valuation, and found flat prices to be overvalued by 35%-70% (depends on approach) in 2008 and by approximately 10% in 2017. These aggregate values also partly confirm our underlying findings, but again, regional differences cannot be validated.

5.4 Discussion of Shortcomings

The main weaknesses of every model are always its underlying assumptions, that are often far from reality. Approaches applied in the thesis also have some important insufficiencies, that were summarized for example by Hlaváček and Komárek (2009) or Plašil and Andrle (2019).

First of all, the results are very sensitive to the choice of variables. There is not a single widely used set and the decision was based on literature in combination with economic intuition and data availability. Problems arise either if an important fundamental determinant is omitted, or if an unimportant one is included. The estimated overvaluation is likely to be higher than actual in the first case, while in the second, problems of overfitting arise (i.e. making the model not very general). Additionally to this, it is very hard to distinguish whether the relationship between factors and flat prices suggested by the model is indeed fundamental, or whether just temporary deviations are observed. Secondly, and probably the most importantly, all models assume that the explanatory variables are always in equilibrium, which appears to be very unrealistic, especially for such a long period. If one of the factors experience

[16] Methodology was already described in Chapter 2. The property market analysis was not published in more recent publication of 2018/2019.

similar deviations during the bubble period as flat prices,[17] the model will likely mistakenly indicate equilibrium. The solution would be either to exclude these bubble periods completely, which would obviously require to know them apriori, or recalculate all fundamentals to their equilibrium values. Lastly, another feature of regression analysis is that it assumes the total sum of the residuals (deviations from equilibrium) to be equal to zero. In other words, it assumes that there is as many undervaluations as overvaluations, which is rarely the case in a sense that according to empirical research, bubbles usually prevail. And since the inclusion of past housing price bubbles in sample is also likely to bias the estimated equilibrium upwards, the actual price bubble is expected to be greater than estimated from models.

Some of the issues above could be addressed by Vector autoregression models (VAR), Vector Error Correction models (VECM), Dynamic stochastic general equilibrium models (DSGE) or, for example, by less traditional Prudential and Valuation indicators.

[17] According to Hlaváček and Komárek, both factors include the same "bubble component".

Chapter 6

Conclusion

This thesis studied fundamental determinants of real estate prices with the main purpose of indicating price bubbles in the regions of the Czech Republic between 2000 and 2017. The main contribution lies in the utilization of balanced panel data set over a long time-period, as well as in deepening the analysis conducted by Hlaváček and Komárek (2009). Additionally, I attempted to increase the robustness of results by conducting two alternative approaches, differing mainly in the way of dealing with nonstationary data and avoiding problems of spurious regression, which were then evaluated simultaneously.

In the preliminary analysis, stationarity tests produced results that are in line with the literature, regardless of frequency of data, with the only exception of real interest rate. Thanks to the approach of first differencing, which was then applied on yearly data, it was confirmed that the prices are mainly driven by demand factors and not by the supply side. As many variables as possible were considered out of which wages, unemployment, natural increase in population and, additionally to similar studies, net migration were all indicated as fundamental price determinants. Mortgages, marriages and divorces turned out less important, but still sufficiently significant, and the vast majority of coefficients had expected signs.

In the second approach, which made use of quarterly data, Westerlund and Kao tests confirmed the cointegrating relationship. In spite of ambiguous results from Pedroni, this allowed to proceed to the estimation of long run equation, in which all considered factors were found to be significant, in other words, this model additionally confirmed the importance of construction prices and real interest rates. In the second stage, thanks to short run equation and its error correction specification, the estimated adjustment term turned out to have required properties for convergence and showed a very slow process of approximately 4.3% adjustment to equilibrium in every quarter.

Since both approaches produced models with desirable characteristics, it was considered appropriate to also utilize them together with simple economic indicators for evaluation of real estate bubbles in individual regions. Price-to-income and price-to-rent ratios indicated the highest risk of overvaluation in Prague and Jihomoravský and recognized price bubbles in periods 2003, 2008 and 2017. Models of house price determinants generally struggled to produce the same results as the yearly model tended to underestimate the deviations in comparison to the quarterly. Nevertheless, both usually confirmed the three main bubble periods already suspected from indicators, however, as they did not occur in all regions at the same time, 4 subgroups were distinguished. Notably, the majority group of regions suggested bubble periods for 2003 and 2008, while only Prague and Jihomoravský seem to currently face a high risk of price overvaluation. A very specific nature of Ústecký was also identified, most likely caused by high levels of unemployment in the early years. Furthermore, models suggested significant undervaluation in current years in regions Liberecký, Karlovarský and Ústecký.

Generally, the estimated degree of overvaluation appeared to be lower compared to the literature, which is likely to be explained by the shortcomings discussed in Section 5.4.

Bibliography

Arellano, M. (1987). Practitioners'corner: Computing robust standard errors for within-groups estimators. *Oxford bulletin of Economics and Statistics*, 49(4):431–434.

Belke, A. and Keil, J. (2018). Fundamental determinants of real estate prices: A panel study of german regions. *International Advances in Economic Research*, 24(1):25–45.

Bill, S. and Ivarsson, M. (2009). "Gimme Shelter"-a long-and short-run analysis of swedish regional house prices.

Cempírek, P. (2014). Housing prices in the czech republic and slovakia: Regional comparison. Bachelor's thesis, Charles University in Prague.

Ćetković, J., Lakić, S., Lazarevska, M., Žarković, M., Vujošević, S., Cvijović, J., and Gogić, M. (2018). Assessment of the real estate market value in the european market by artificial neural networks application. *Complexity*, 2018.

CNB (2017/2018). Financial stability report 2017/2018. Available at: `https://www.cnb.cz/en/financial-stability/fs-reports/financial-stability-report-2017-2018/`.

CZSO (2018). Ceny sledovaných druhů nemovitostí - 2015 až 2017. Available at: `https://www.czso.cz/csu/czso/`.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431.

Égert, B. and Mihaljek, D. (2007). Determinants of house prices in central and eastern europe. *Comparative economic studies*, 49(3):367–388.

Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276.

EViews (2019). *EViews User's Guide*. Available at: `http://www.eviews.com/help/helpintro.html#page/content/preface.html`.

Granger, C. W. and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of econometrics*, 2(2):111–120.

Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *The Econometrics Journal*, 3(2):148–161.

Himmelberg, C., Mayer, C., and Sinai, T. (2005). Assessing high house prices: Bubbles, fundamentals and misperceptions. *Journal of Economic Perspectives*, 19(4):67–92.

Hlaváček, M. and Hejlová, H. (2014). A comprehensive method for house price sustainability assessment. *Financial Stability Report*, 2015:121–130.

Hlaváček, M. and Komárek, L. (2009). Housing price bubbles and their determinants in the czech republic and its regions. *CNB WP*, 12:2009.

Hurlin, C. and Mignon, V. (2007). Second generation panel unit root tests.

Im, K. S., Pesaran, M. H., and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of econometrics*, 115(1):53–74.

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of econometrics*, 90(1):1–44.

Levin, A., Lin, C.-F., and Chu, C.-S. J. (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of econometrics*, 108(1):1–24.

Loayza, N. and Ranciere, R. (2004). *Financial development, financial fragility, and growth*. The World Bank.

Mikhed, V. and Zemčík, P. (2009). Do house prices reflect fundamentals? aggregate and panel data evidence. *Journal of Housing Economics*, 18(2):140–149.

Newey, W. K. and West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.

Njoroge, P. K., Aduda, J. A., and Mugo, C. (2018). Investigating the existence of a bubble in the kenyan real estate market. *International Journal of Data Science and Analysis*, 4(5):89.

Pedroni, P. (2004). Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the ppp hypothesis. *Econometric theory*, 20(3):597–625.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of applied econometrics*, 22(2):265–312.

Phillips, P. C. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.

Plašil, M. and Andrle, M. (2019). Assessing house price sustainability. *Czech National Bank*.

Rae, D., Van den Noord, P., et al. (2006). Ireland's housing boom: What has driven it and have prices overshot? Technical report, OECD Publishing.

Schwert, G. (1989). Tests for unit roots: A monte carlo investigation. *Journal of Business & Economic Statistics*, 7(2):147–59.

Stohldreier, M.-T. (2012). The determinants of house prices in chinese cities. *Mater Thesis of University of Zurich*.

Sunega, P., Lux, M., and Zemčík, P. (2014). Housing price volatility and econometrics. *Crit. Hous. Anal*, 1(1).

Tsatsaronis, K. and Zhu, H. (2004). What drives housing price dynamics: cross-country evidence. *BIS Quarterly Review, March*.

Čadil, J. (2009). Housing price bubble analysis–case of the czech republic. *Prague Economic Papers*, 1:38–47.

Westerlund, J. (2007). Testing for error correction in panel data. *Oxford Bulletin of Economics and statistics*, 69(6):709–748.

Wooldridge, J. M. (2013). Introductory econometrics: a modern approach 5th edition. *Mason, OH: South-Western*.

Zemčík, P. (2011). Is there a real estate bubble in the czech republic? *Forthcoming in the Czech Journal of Economics and Finance*.

Appendix

Table 1: Yearly data - Panel stationarity tests

Variable	**Levels**				**Differences**				**Result**
	LLC	**Hadri**	**IPS**	**CIPS**	**LLC**	**Hadri**	**IPS**	**CIPS**	
rprice	-2,453	7,059	-1,423	-1,069	-11,199	-0,610	-7,306	-2,280	I(1)
	0,007	0,000	0,077	0,100	0,000	0,729	0,000	0,059	
rwage	1,973	9,737	2,067	-1,701	-1,910	1,020	-2,121	-3,285	I(1)
	0,976	0,000	0,981	0,100	0,028	0,154	0,017	0,010	
unemployment	1,080	7,846	-0,231	-1,865	-10,812	-0,012	-7,629	-3,292	I(1)
	0,860	0,000	0,408	0,100	0,000	0,505	0,000	0,010	
natural	-4,178	4,766	-3,090	-2,990	-6,282	0,520	-6,732	-4,076	I(0)
	0,000	0,000	0,001	0,010	0,000	0,301	0,000	0,010	
net migration	-3,654	-0,121	-2,671	-2,500	-15,147	-0,489	-11,767	-3,692	I(0)
	0,000	0,548	0,004	0,019	0,000	0,688	0,000	0,010	
marriages	-1,813	5,110	-1,810	-2,107	-10,115	0,747	-7,591	-2,939	I(0)
	0,035	0,000	0,035	0,100	0,000	0,227	0,000	0,010	
divorces	-1,731	8,056	3,498	-2,318	-15,061	2,987	-13,055	-3,433	I(0)
	0,042	0,000	1,000	0,048	0,000	0,001	0,000	0,010	
new flats	-2,413	1,273	-2,352	-2,116	-18,069	0,409	-15,079	-2,934	I(0)
	0,008	0,101	0,009	0,100	0,000	0,341	0,000	0,010	
activity	-2,844	5,064	-2,125	-2,034	-9,624	4,047	-7,506	-3,206	I(0)
	0,002	0,000	0,017	0,100	0,000	0,000	0,000	0,010	

Note: Statistic on top, p-value on bottom. All tests except CIPS (in R) were performed in EViews. Only intercept was chosen as deterministic component. Lag length selection based on SIC. Coloured values indicate stationarity at 5% significance level (reversed null hypothesis for Hadri). Author's calculations.

Table 2: Yearly data - Time series stationarity tests

Variable	Levels		Differences		Result
	ADF	PP	ADF	PP	
rinterest	-3,603	-3,603	-5,431	-15,956	I(0)
	0,017	0,017	0,001	0,000	
rmortgage	-1,091	-1,146	-3,437	-2,719	I(1)
	0,694	0,672	0,026	0,093	
rconstruction	-1,078	-0,676	-2,036	-5,144	I(1)
	0,697	0,827	0,270	0,002	

Note: Statistic on top, p-value on bottom. Tests performed in EViews. Only intercept was chosen as deterministic component. Lag length selection based on SIC. Coloured values indicate stationarity at 5% significance level. Author's calculations.

Table 3: Hausman test results

	Model 1	Model 2	ECM
stat	51,10	20,18	12,248
p-val	0,000	0,043	0,03154

Note: Null hypothesis: RE is inconsistent. Author's calculations.

Table 4: Quarterly data - Panel stationarity tests

Variable	Levels				Differences				Result
	LLC	Hadri	IPS	CIPS	LLC	Hadri	IPS	CIPS	
rprice	-3,609	11,698	-0,761	-1,142	-11,534	-1,135	-10,874	-3,347	I(1)
	0,000	0,000	0,223	0,100	0,000	0,872	0,000	0,010	
rwage	2,724	22,192	2,066	-1,795	-33,434	1,417	-29,308	-3,835	I(1)
	0,997	0,000	0,981	0,100	0,000	0,078	0,000	0,010	
unemployment	2,087	12,410	0,519	-2,035	-31,791	-1,188	-29,497	-4,089	I(1)
	0,982	0,000	0,698	0,100	0,000	0,883	0,000	0,010	

Note: Statistic on top, p-value on bottom. All tests except CIPS (in R) were performed in EViews. Only intercept was chosen as deterministic component. Lag length selection based on SIC. Coloured values indicate stationarity at 5% significance level (reversed null hypothesis for Hadri). Author's calculations.

Table 5: Quarterly data - Time series stationarity tests

Variable	Levels ADF	Levels PP	Differences ADF	Differences PP	Result
rinterest	-1,953	-2,701	-6,369	-6,851	I(1)
	0,307	0,079	0,000	0,000	
rconstruction	-1,232	-0,323	-2,154	-7,818	I(1)
	0,656	0,915	0,225	0,000	

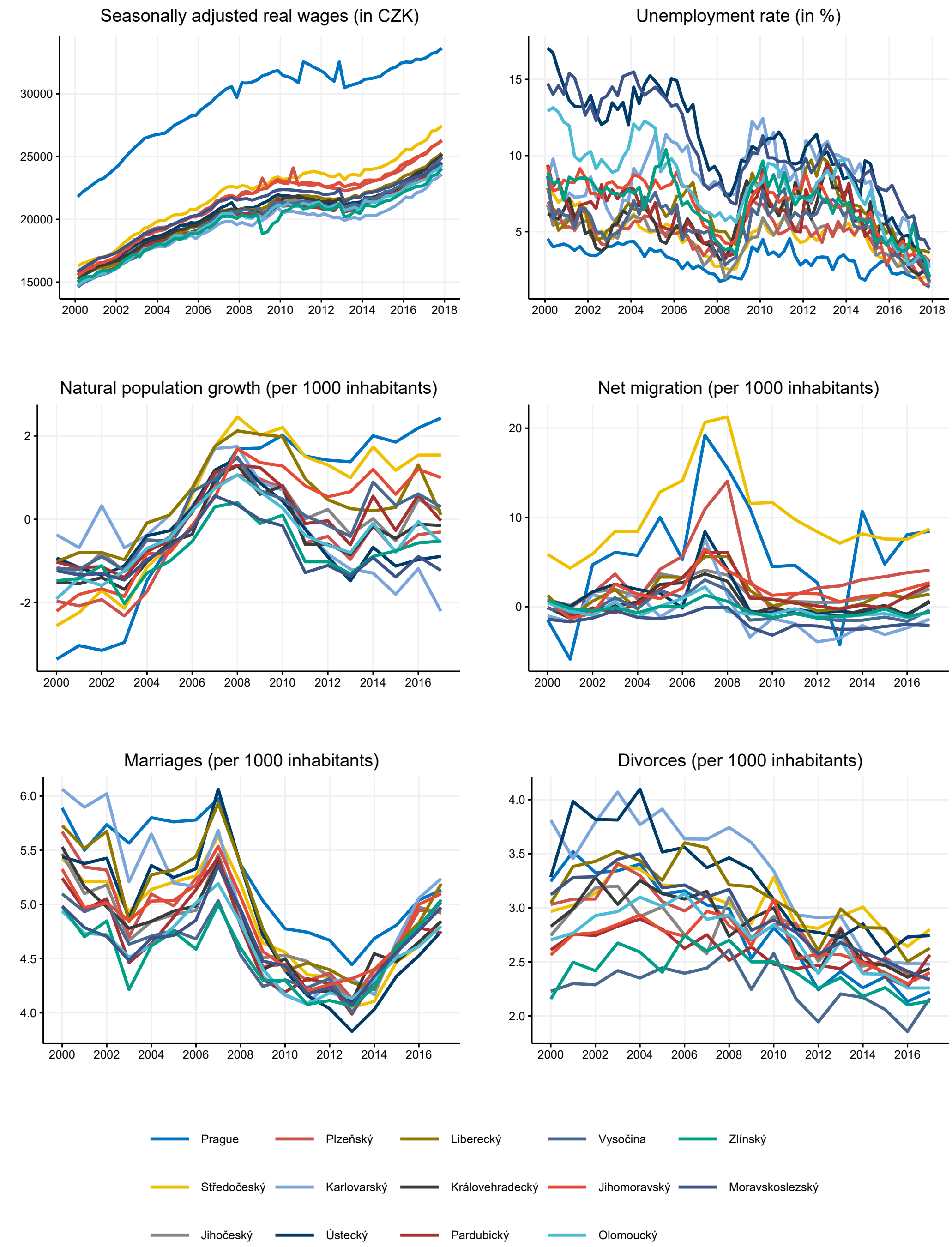
Note: Statistic on top, p-value on bottom. Tests performed in EViews. Only intercept was chosen as deterministic component. Lag length selection based on SIC. Coloured values indicate stationarity at 5% significance level. Author's calculations.

Table 6: Quarterly analysis: Short run results

	Dependent variable: Δ rprice	Expected sign
ECM	−0.043***	−
	(0.009)	
Δ rwage	0.356***	+
	(0.063)	
Δ unemployment	−112.588***	−
	(23.440)	
Δ rconstruction	189.095***	+
	(34.886)	
Δ rinterest	−121.711***	−
	(15.704)	
Observations	970	
R^2	0.133	
Adjusted R^2	0.116	

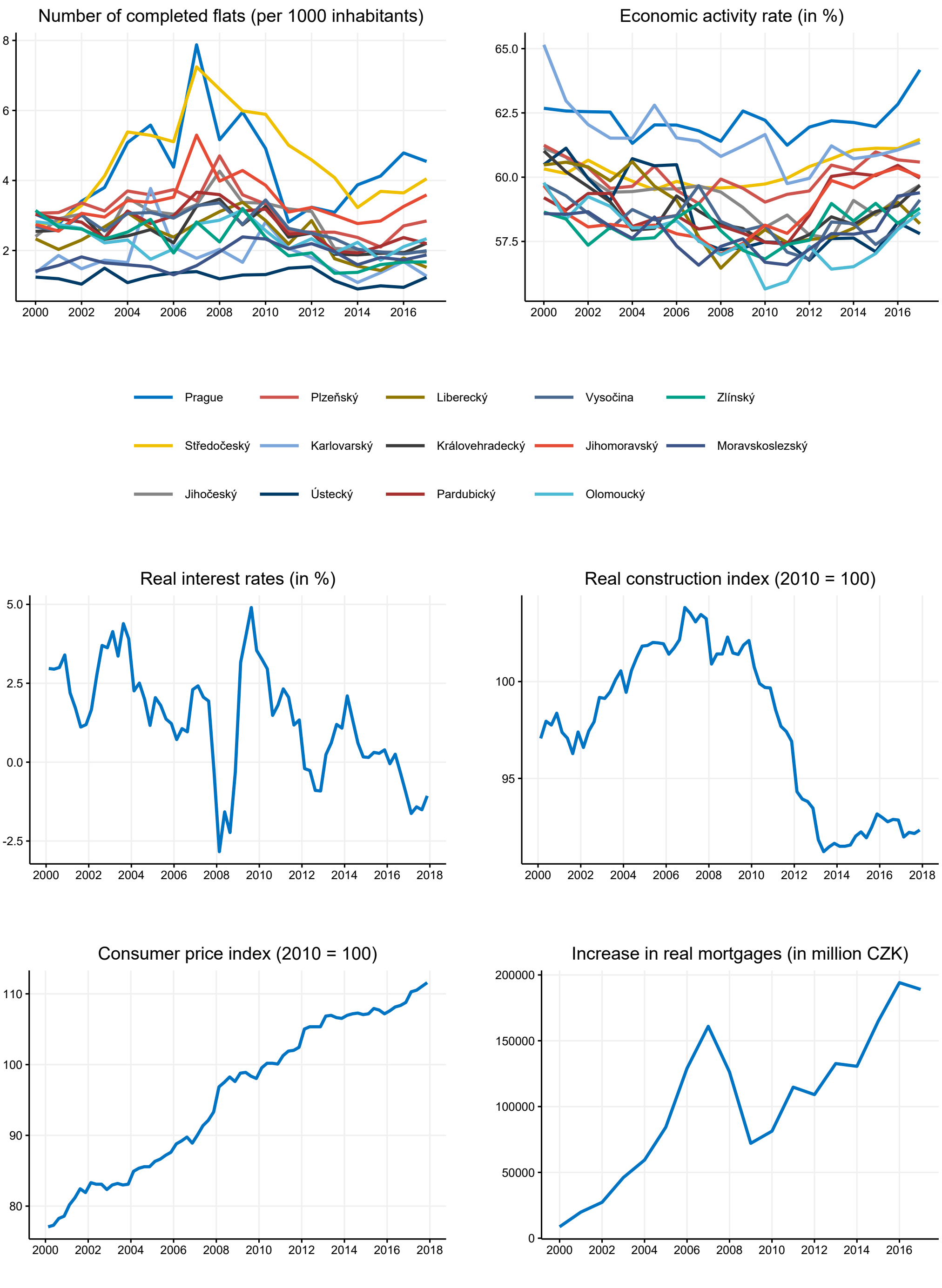
Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$ *. Estimated by FE. Arellano robust errors. rwage seasonally adjusted. Author's calculations.*

Figure 1: Data set part I



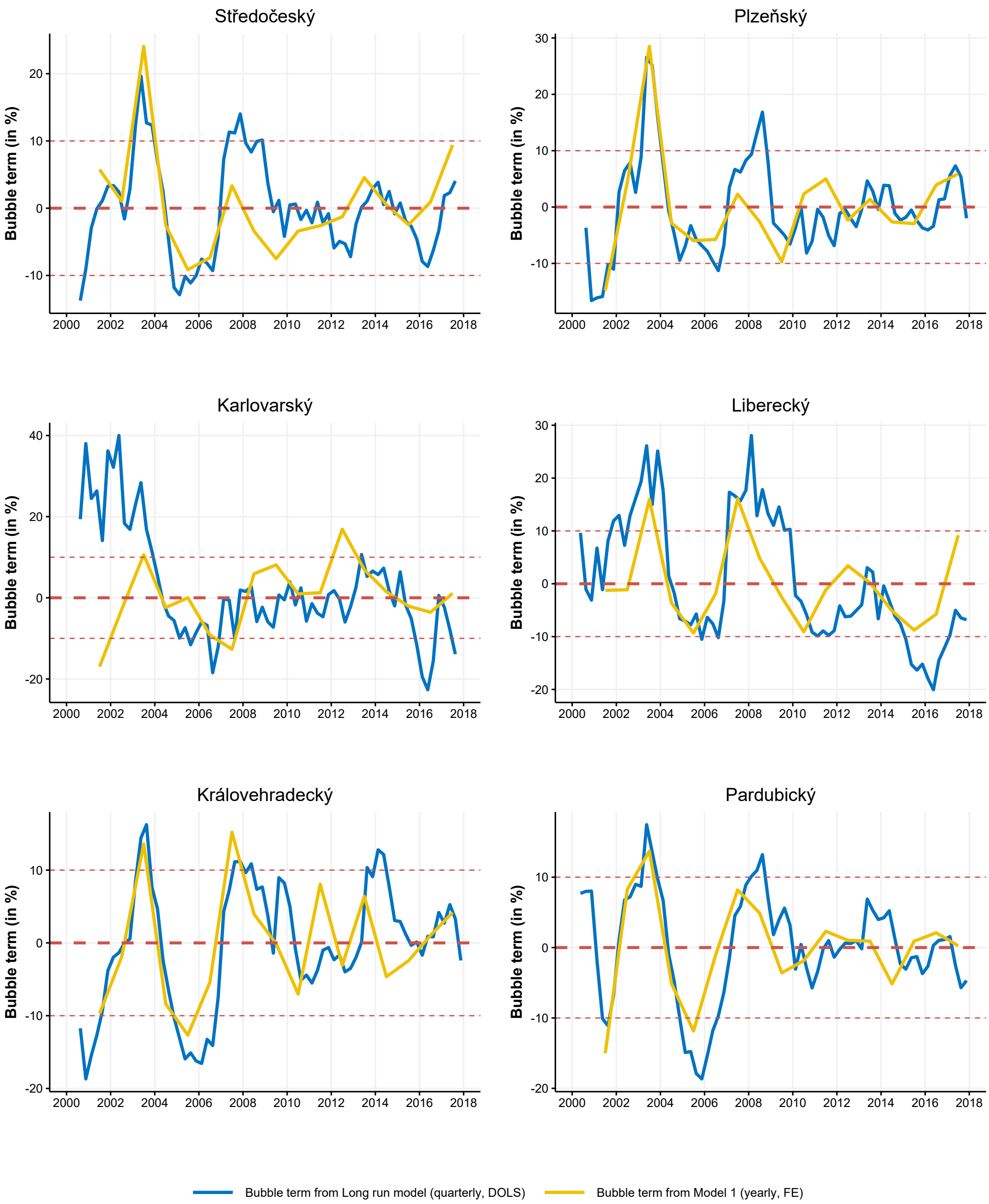
Note: real variables are 2010 prices, graphs generated in R

Figure 2: Data set part II



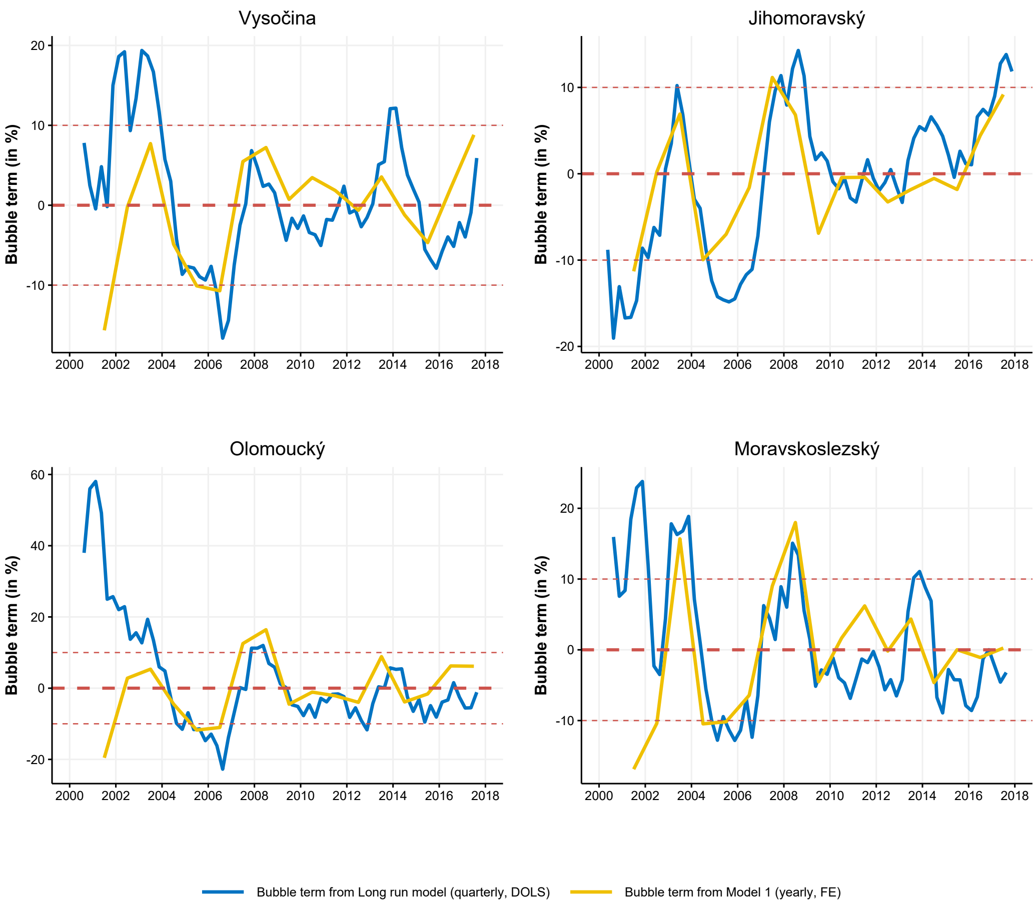
Note: real variables are 2010 prices, graphs generated in R

Figure 3: Real Estate Bubbles in other regions (Part I)



Note: Values of Bubble term evenly distributed. Graphs generated in R.

Figure 4: Real Estate Bubbles in other regions (Part II)



Note: Values of Bubble term evenly distributed. Graphs generated in R.