CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Determinants of Real Estate Prices in the United States

Bachelor's Thesis

Author of the Thesis: **Michaela Hronová**

Supervisor: **Mgr. Roman Kalabiška**

Year of the defense: **2022**

Declaration

1. I hereby declare that I have compiled this thesis using the listed literature and resources only.
2. I hereby declare that my thesis has not been used to gain any other academic title.
3. I fully agree to my work being used for study and scientific purposes.

In Prague on 28th of July 2022 Michaela Hronová

References

HRONOVÁ, Michaela. Determinants of Real Estate Prices in the United States. Praha, 2022. 82 pages. Bachelor's thesis (Bc). Charles University, Faculty of Social Sciences, Institute of Economic Studies. Supervisor Mgr. Roman Kalabiška

Length of the Thesis: 113 043 characters

Abstract

This thesis aims to examine the fundamental determinants of real estate prices in the United States using State level quarterly data over a time period of 21 years. Determinants from both the supply and demand sides are used for the analysis. After the determination of non-stationary variables and further confirmation of a single cointegrated relationship between home values and personal income variables, the ARDL model was implemented. With the use of the ARDL model, the effect of selected variables on home values was analyzed as well as the error correction term, which estimates the speed of adjustment towards the equilibrium following a change in other variables. From the supply factors, the regression revealed a positive effect of building permits which were significant in all 9 divisions, and a negative effect of homeowner vacancy rates which were significant in 6 divisions. From the demand factors, personal income per capita was identified as the main determinant together with unemployment, population growth, and marriage rates.

Keywords

Housing market, home values, cointegration, panel regression

Abstrakt

Cílem této práce je prozkoumat základní determinanty cen nemovitostí ve Spojených státech za pomocí čtvrtletních dat na státní úrovni po dobu 21 let. K analýze jsou použity determinanty jak ze strany nabídky, tak i poptávky. Po stanovení nestacionárních proměnných a po následném potvrzení kointegrovaného vztahu mezi hodnotou domů a osobním příjmem, byl použit model ARDL. S využitím modelu ARDL byl analyzován vliv vybraných proměnných na hodnoty domů a následně také člen korekce chyb, který odhaduje rychlost přizpůsobení k dlouhodobé rovnováze po změně ostatních proměnných. Z nabídkových faktorů regrese odhalila pozitivní vliv stavebních povolení, která byla shledána významnými ve všech 9 divizích a negativní vliv míry obsazenosti domů, která byla významná v 6 divizích. Z poptávkových faktorů byl identifikován jako hlavní determinant osobní příjem spolu s nezaměstnaností, populačním růstem a sňatečností.

Klíčová slova

Nemovitostní trh, hodnoty domů, kointegrace, panelová regrese

Název práce

Determinanty cen nemovitostí ve Spojených Státech

Table of Contents

List of tables

1 Introduction

The law of supply and demand may be used to explain real estate prices, but it is difficult to achieve equilibrium since the supply reacts to changes in demand slowly. When the demand for more housing units is higher than the current availability on a market, home sellers increase the prices to ration the available supply. This price increase and the demand which cannot be satisfied by the already existing stock of housing units act like an incentive for investors and developers to build new houses. However, developers will produce these additional units with a lag due to construction delays and the fact that new land needs to be made available, which is also a time-consuming task. (Glickman, 2014) As a result of long construction lags, the supply side of the real estate market cannot respond quickly to the increasing demand. Hence, housing price booms may emerge due to the initial lack of supply and later reversion due to excess supply once the demand drops. (Belke and Wiedmann, 2005)

In the United States, the ownership of property is frequently a significant source of household wealth, and the construction industry generates a sizable number of jobs. Therefore, housing accounts for a substantial portion of all economic activity, and changes in the housing sector can have a wider impact on the economy. The estimation of home prices is a major concern for buyers, investors, and the financial sector. Thus, many papers discuss the effect of various economic variables on house prices.

This thesis aims to extend the regression analysis from previous studies by the use of longer time series and by the inclusion of multifarious factors from the supply and demand side. Furthermore, this thesis attempts to determine to what extent the house prices in the United States and divisions are driven by fundamentals like personal income per capita, unemployment, and others. Another objective is to determine the speed of adjustment toward the long-term equilibrium. A panel dataset of 49 states plus the District of Columbia over the time period of 21 years quarterly from the first quarter of the year 2000 to the fourth quarter of the year 2020 was used for the analysis. This dataset allows for the robust evaluation of long-term macroeconomic effects.

The structure of the thesis presented is organized as follows. The following chapter describes the real estate market in the United States, its history as well as current events. Subsequently, multifarious determinants of house prices are described in detail and the

possible effects on property prices are deliberated. Chapter 4 provides an overview of the empirical literature and studies on comparable topics. Chapter 5 explains the selection of data and further describes the sources and frequencies of the selected data. It continues with the preliminary analysis where multiple stationarity tests are applied and based on the results, cointegration tests are conducted. Subsequently, the proper model was chosen, and the analysis was performed. Finally, the results of the regression were presented and discussed in chapter 6.

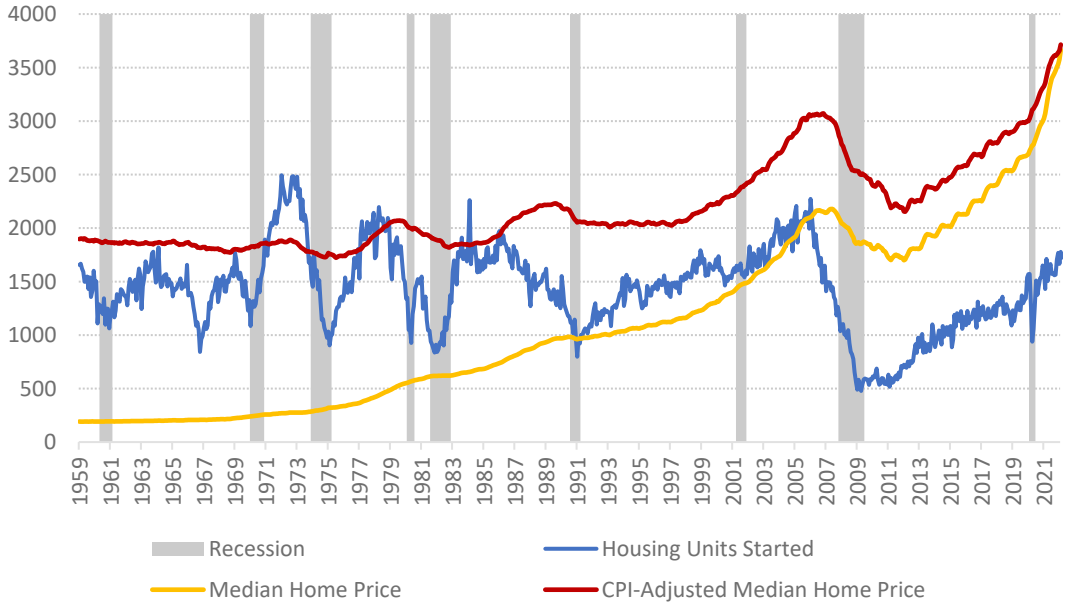
2 Real estate market in the United States

The real estate industry is a major contributor to the U.S. economy and a capital-intensive endeavor. The property markets' economic health and their changes can have a significant impact on the overall economy. With an increase in the housing value, homeowners spend more money than they normally would because they have higher confidence in the economy, see a possible greater rental income, and can borrow against the increased equity. A fall in pricing has the opposite effect. In the United States, personal consumption expenditures (PCE) represent 68.5 % of the total GDP, therefore, changes in housing wealth can have an immense impact on economic growth. Because many economies throughout the world are so intertwined with the U.S. economy, changes in the U.S. housing market have reverberations all over the world. This reality was highlighted in 2007, when a local real estate crisis in American markets grew to global dimensions, threatening to put the global financial system to a standstill. (Juneja, 2021)

2.1 History of the U.S. housing market

In each state, the nation´s economy and the overall health of the economy fluctuate over time. In the U.S. there were several declines over the years, each different in case of severity. The two most well-known are the Great Depression in the 1930s and the Great Recession in 2008. The median home price, which is displayed in hundreds of dollars in Figure 2.1 demonstrates that from 1963 the prices fairly steadily increased. However, in almost every decade there was a decline for various types of reasons. These recessions are indicated by the shaded areas. Throughout each recession there was a significant drop in the housing units started, which affected the housing supply, thus increasing the price of real estate, which continued to grow until The Housing Crisis.

Figure 2.1 Average home prices and housing units started



Some people place the start of the housing bubble in 1997. Since then, the national housing prices marched upwards until 2006. The Case-Shiller national house price index more than doubled between 1995 and 2005. And since the middle of 2006, real estate values have been plummeting. Others believe that the start was in 2001-2002 because of the significantly lowered short-term interest rate by the Federal Reserve. Levitin and Wachter believe that the bubble started in 2004 (possibly 2003) because until then the price increase of houses was explained by the fundamentals, which were interest rates and the cost of homeownership relative to renting. Based on inflation-adjusted housing prices they examined that the prices did not pass their previous peak level until 2000. In 2000 housing prices started to increase much faster than rental prices thus tied to the lower interest rates in 2001-2003, the cost of homeownership fell (Levitin and Wachter, 2010). In the early 2000´s the mortgage lending standards loosened up, so it was possible for people to obtain a mortgage loan even if they had a really low credit score. Even then they qualified for subprime loans, which were low-cost and low-down-payment mortgages. These subprime loans were then put together with high-quality, low-interest loans and sold on large scales to investors. Demand for these packaged loans grew and so did the capital in U.S. mortgage-backed securities. Because of these loose lending standards, banks allowed practically anyone to borrow money without even considering their ability to pay back or regarding the debt-to-income ratios. Adjustable-rate mortgages (ARM) allowed the borrower to make much lower initial payments but after two or three years the ARMs

would reset and then the payments could vary each month and be much larger than the initial payments. In the beginning, these mortgages seemed risk-free, and the borrower could end up with positive equity as they could sell the property with a profit if they were not able to pay the higher payments. But with the downturn of the housing market, this would leave the borrowers/owners of the property in an impossible situation as they would not be able to sell, thus ending with negative equity. By March 2007 it became evident that the subprime market was in trouble. Millions of homeowners defaulted because they were not able to pay back the payments when they were due, and many builders and developers went bankrupt. The prices of homes declined, leaving the homeowners in a situation where they were no longer able to sell their homes for a profit and they had to lose their houses to foreclosure and often file for bankruptcy.

In September of 2008 when Lehman Brothers collapsed, the housing market crashed and with it so did the stock market, thus leaving the U.S. and the entire world in the Great Recession. During the Recession, house prices dropped by approximately 30 % and it took over a decade for the housing market to recover. By 2012 homebuyers and hedge funds from Wall Street started buying foreclosures and home values increased. Based on Dodd-Frank Wall Street Reform and Consumer Protection Act the lending standards were tightened to prevent another housing crisis. (Fettke, 2020)

2.2 Covid-19 pandemic

Economic activity was halted in March 2020, resulting in an increase in unemployment as a result of lockdown orders. COVID-19 caused a negative shock to household earnings, both present, and expected incomes, as a result of widespread layoffs, putting downward pressure on property prices due to lower demand. There were differences between household types and between the two sides of the income distribution. Unemployment or wage losses were more likely to affect low-income households or minorities. Workers in small companies were more likely to be laid off which could have resulted in defaulting on mortgages or property owners losing their rent. (Balemi, Füss and Weigand, 2021) Low-interest rates and social separation boosted the desire for privately owned dwellings, even though families cut overall expenditure and increased savings due to increased economic uncertainty. Mostly because of precautionary motives or liquidity preferences. Furthermore, the shift in the workplace toward home office (remote employment), as well as social isolation, raised the demand for more privacy and living

space. To reduce these negative shifts in the supply and demand of goods, governments used a wide range of monetary and fiscal policy instruments, including unemployment and financial assistance programs, helicopter money, short-term working benefits, mortgage forbearances, rent moratoriums for tenants, and reductions in the federal funds rate. (Balemi, Füss and Weigand, 2021)

After The Fed's monetary easing the year-on-year growth rate of median house price per square foot advanced quicker from May to September than in any four-months period before the global financial crisis of 2007-2009. One of the reasons for higher prices was the decreasing trend of the housing supply since mid-2019, which was further accelerated by COVID-19. From 2001 to 2020, U.S. builders added only 1.23 million new housing units per year on average, which created a shortage of 7 million units on the market that would be needed to fulfill household formation growth and replace units that were aging or destroyed by natural disasters. Quarantine orders stopped national homebuilders from securing building permits. In addition, labor shortage and lack of ready-to-build land and materials further slowed down the pace of construction growth, thus contributing to the country's already depleted inventory. At the same time, because of the anxiety and uncertainty caused by COVID-19, homeowners removed their listings from the market. As a result, the housing market in the United States lacks sufficient inventory. According to Zillow´s data, during the pandemic from January 2020 to January 2022, the average home price increased by 23.6 %.

In April 2020, movements in home prices, demand, and supply were similar across the country, in urban, suburban, and rural areas, indicating that rising housing prices were a nationwide phenomenon caused by monetary easing. Lower mortgage rates induced housing demand because they encouraged people, particularly wealthy households with the financial means to make several down payments, to take out another mortgage and buy a property for speculative purposes. Zhao's findings revealed that most middle-class people would just refinance their existing mortgages to take advantage of the low rates. (Zhao, 2020)

2.3 Differences in real estate market across States

When it comes to ethnicity, religion, and demography, the United States is one of the most heterogeneous countries in the world. Prices of houses vastly differ from state to state, from county to county, and even from city to city. In this chapter, some of the reasons behind the price differences will be explained.

2.3.1 Population density

The population density in the United States in 2022 is 34 people per square kilometer, where the population density is calculated as if the population were to be evenly distributed across the land area. However, most people tend to cluster in cities, which thus have a much higher density. In the U.S., 97 % of the land mass is considered to be rural areas but only 17.34 % of the total population lives there, which makes the remaining 82.66 % of the population live in 3 % of the urban areas. (El Nasser, 2018) (Statista, 2022b) Density levels vary significantly from city to city, and even within a single city. Major metropolitan areas in the states of California, New York, New Jersey, and Connecticut, among others have grown so popular that they are now experiencing overcrowding. For example, the New York-Newark-Jersey City Metro area has one of the highest population densities with more than 1,110 inhabitants per square kilometer, second highest would be the Los Angeles-Long Beach-Anaheim metropolitan area in California which has approximately 1,050 people per square kilometer. (U.S. Census Bureau, 2020) (U.S. Census Bureau, 2016)

Using population densities is most useful when comparing small areas like neighborhoods. By using state population densities, we can compare the settlement intensity across states. Looking at the states, New Jersey has the highest population density with 486 people per square kilometer. It is the 5th smallest state when considering the land area but the 11th most populous state. On the contrary, Alaska is the largest state in the U.S. with a total area of 1,723,337 km^2 and only 733,391 inhabitants, making it the least populous state with only 2 people per km^2. (U.S. Census Bureau QuickFacts, 2021)

Based on Census data, the most common reason for moving is finding more affordable housing (55 %), and the second reason is work-related moving, either relocation or finding a more prospering job (18 %). Counties in the south and west regions experience the highest population increase. In 2018, states like Florida, Texas, and Utah registered the highest positive domestic migration. The reason behind it is that the top ten fastest-growing

metro areas can be found within these states. In a year, more than 560 thousand people moved to Florida. On the contrary, California had the highest number of out-movers, approximately 660 thousand people. These movers usually relocate to areas close to California, on the west coast. These states are attractive for a few reasons. Firstly, the Pacific West has thriving creative marketing and technology businesses, particularly in Portland and Seattle. Secondly, the low cost of living and access to green spaces and outdoor activities attract retirees as well as families. The increased demand thus drives population growth upwards.

Cities and their real estate markets can be divided into 3 Tiers, based on the stage of development of their markets, demographics, and job growth. Tier I. may consist of cities like New York, Los Angeles, Chicago, Boston, San Francisco, and Washington D.C, these cities are considered to be the centers of long-established commerce and population. In Tier II we would find Salt Lake City, Miami, Seattle, Atlanta, Dallas, Austin, and others. These cities offer similar amenities to the ones from Tier I., but they have lower population densities. Tier III. cities have emerging markets with steady but restricted job growth, populations of usually less than one million people, and a mix of economic drivers. (Chen, 2020) (Balint, 2018)

The average monthly rent in Manhattan, which is considered to be in Tier I., was $4,100 in 2018, the yearly rent thus represented 59 % of the median yearly income of a household, where the median income was $83,500. Similarly, at the beginning of 2022, the average rent in LA county was $2,661, which makes up approximately 45 % of the median yearly income of a household. Lately, this is the reason Americans move to second-tier cities or suburbs, where the housing is more affordable. As an example, in the second-tier metro, Maricopa, the average value of a home is $280,000 which makes up 24 % of the yearly median income of a household. (Hiscock, 2016)

2.3.2 Property taxes

Another difference is property taxes which can vary significantly from state to state. The nationwide effective property tax rate is 1.1 % of the average home value, but the value of property tax is determined by each county and thus can differ significantly even between two neighboring counties. Overall Hawaii is the most expensive state when looking at the average home price, which was $679,000 at the beginning of 2022, however, their average effective property tax rate per year is only 0.28 %, where this value is

calculated as a yearly median percentage of property's assessed fair market value. On the other hand, New Jersey has the highest property tax rate at 2.47 %. When comparing the annual property tax in these two states, with the assessed home value of $250,000 one would pay $6,175 in New Jersey, however for the same valued home only $700 in Hawaii. The collected taxes are then used by the local governing authority to pay for water and sewer improvements, road and highway development, as well as law enforcement, fire protection, libraries, education, and other community services. (Fritts, 2022) The education factor is one of the main determinants of the property tax rate due to the importance of school ratings which determine the school tax rates in the neighborhood. When trying to grow the population some cities use the lowering of tax rates to attract more residents, whereas communities that have been built over time may have to charge more to cover the expenses for the already built infrastructure.

2.3.3 Transportation

For most Americans owning a car is a necessity because of the lack of public transportation. However, for individuals living in large metropolitan areas which offer reliable public transportation, it does not need to be the case. Owning a car comes with certain advantages as well as disadvantages. Advantages being the flexibility and independence which people owning a car have, traveling at any time and to any place they need without having to rely on public transportation. On the other hand, a car requires additional expenses such as insurance, fuel, and paying for a parking space. In 2020 the average number of vehicles per licensed driver was 1.21 or 832 vehicles per 1,000 inhabitants of the US. When comparing this number across states, one can see that the numbers differ vastly. In states where there are big cities that provide public transportation the vehicle ownership rates are significantly smaller. As an example, New York City is considered to have the best public transport and in the state of New York in 2017, the number of vehicles per 1,000 inhabitants was 539. On the contrary, in rural states, where many households practice agriculture production the vehicle ownership rates are as high as 1,140 vehicles per 1,000 inhabitants. This number corresponds with the state of Wyoming, where the vehicle ownership rate was the highest in 2017, however, similar numbers were seen in states like Idaho and Montana. (Pariona, 2017)

Data from the Bureau of Labor Statistics show that transportation is the second biggest component of the household´s annual average expenditures right after housing, where in 2020 housing accounted for nearly 35 % of the average expenditures and transportation

accounted for 16 %. (Bureau of Labor Statistics, 2021) This transportation component includes vehicle purchases, gasoline, other fuels, motor oil, and public transportation. However, even added costs can arise from owning a car like insurance payments and the need for parking spaces. These expenses may be reflected in house prices, because with higher expenses the available funds to purchase a house or to repay the mortgage loan decreases.

3 Housing price determinants

The law of demand and supply may be used to explain real estate prices in equilibrium. However, due to the unique characteristics of the real estate market and other inefficiencies, which will be further explained in this chapter, it is not always simple to maintain the balance between supply and demand. Supply and demand factors of the housing market and potential problems connected to it will be mentioned as well.

3.1 Unique characteristics of the housing market

The real estate market differs from the common commodities market, because of the unique properties which the real estate market has. These unique characteristics need to be accommodated in order to apply the supply and demand analysis.

The first unique characteristic is the **durability** of real estate. Houses are built to last for years, decades, or even centuries, which is why real estate is considered to be a durable investment with a long-term store of value. Renovation can further elongate their useful life and the land underneath is fairly indestructible. The vast majority of the housing supply consists of already existing houses. Close to 142.5 million housing units were available at the beginning of 2022 and approximately 1.2 million new houses enter the market each year. In the U.S., the median age of homes is 37 years, depending on the area. The area where the median age of homes is the highest in the Northeast region with a median age of 58 years. On the other hand, houses in areas like Texas, New Mexico, Arizona, or Nevada are considered to be the youngest, usually 20 to 25 years old. (Jones, 2021) Around 3,000 homes can be found in the U.S. that were built 251 to 320 years ago, which makes them 75 years older than the United States itself. (Statista, 2021)

The next unique physical characteristic is the **immobility** of real estate and land. The fixity of the space means that market adjustments occur by people moving to the property, rather than the goods moving to them. That is why the location of the property has an immense impact on the price and demand. The immobility of real estate is connected to transportation discussed in Chapter 2.3.3. The proximity of real estate to public transportation can have a positive impact on house prices. However private transportation comes with additional costs and an increase in these costs can negatively affect house prices. Based on the location another factor can be the property tax discussed in Chapter 2.3.2, which differs based on the county or even neighborhood.

This next characteristic goes hand in hand with the immobility of real estate and that is the **heterogeneity** of real estate. Every real estate is unique in terms of construction, financing, condition and mainly location, the infrastructure of the neighborhood (for example schools, hospitals, shops), neighborhood status, and the commute to your employment. Even two apartments in the same building do not necessarily have to have the same value because even the orientation of the apartment or the floor number can have a large effect on the price. Adding amenities like a terrace, pool, or parking garage can change the property's value, thus assessing the value and comparing the properties can get difficult.

The supply tends to be **inelastic**, due to existing stock, where the decision to sell is up to the current owner. According to the Total Housing Inventory for the United States, in the first quarter of 2022, the whole U.S. had approximately 142.5 million housing units, out of which 90 % were occupied either by homeowners or renters and 15 million housing units were considered to be vacant. However, the vacant units can be further separated into two categories, seasonally vacant and year-round vacant. Another subcategory under the year-round category is a large category called "held off market" which includes units for occasional use, units occupied by persons with usual residence elsewhere, and units vacant for other reasons, for example, foreclosures, renovations, and repairs, abandoned units or units possibly to be demolished and other reasons. Overall, there are 6.977 million housing units in this category and 3.835 million housing units are considered to be seasonally vacant. Thus, only the remaining 4.325 million housing units are for rent or sale. (Housing Vacancies and Homeownership, 2022)

The supply of new properties often adjusts to changes in demand slowly because of **construction lag** which leaves the equilibrium not stable nor efficient in the short run. As a result, home price bubbles may occur due to a lack of supply at first and then reverse due to excess supply as demand falls. (Belke and Wiedmann, 2005) The average amount of time it takes from authorization of residential buildings to start is more than 2 months for buildings with two or more units. Then the construction itself, of a building with 1 unit, takes on average 7.2 months and a building with 2 or more units 15.4 months. (U.S. Census Bureau, 2022a) Another delay can arise even when obtaining permits or securing financing. Lack of information on the real estate market, their complexity, and availability can play a role as well. Extensive research can be costly and time-consuming. In addition,

long-term rental contracts can prevent demand, supply, and prices from adjusting quickly, because these contracts are often enclosed for 3 to 10 years.

The last unique feature is the **transaction costs**, which are relatively significant when it comes to real estate. Firstly, the best-known cost is the *real estate agent fee*. It is a cost that is charged by the real estate agent, a person whose job is to mediate the sale between the seller and the buyer and to help with the negotiations so that both parties are content. For that, they charge a fixed percentage of the transaction value from both of the participating sides. The next commonly known cost is the *search cost*, which consists of an advertisement of the property, or a credit check of the potential renter or buyer. Nowadays, advertisement costs can be minimized by using online platforms. Purchasing a property requires extensive paperwork, and both parties want to ensure that the deal is properly executed. Hence both the parties have their attorneys and lawyers engaged in the process. These would be covered by the *legal and administrative costs*. Examples of administrative costs can be the transfer of title for utilities such as water, electricity, cable, etc. *Statutory costs* are costs usually added by the government. In some of the states in the U.S., stamp duty is levied on the transaction for the transaction to be considered legally valid. The government can also charge a fee for the conveyancing of the property title. Lastly, *financing costs* that are tied to mortgage payments, a fee for a creditor for valuation of the asset, transaction charges that sometimes need to be paid to the lender as well as a processing fee for the mortgage agreement. (The Industry Insights, 2021)

3.2 Supply side

The supply of new housing depends mainly on the production of developers and their profitability. Factors that determine the profitability of a project include the availability and the cost of factors of production, expectations regarding future demand for real estate, and perceived market risk. To proxy the variable of the supply of new housing variable number of building permits per 1,000 inhabitants will be used in the analysis in Chapter 5.

For production, developers use several types of inputs: primarily land, where land is a fixed characteristic of a certain region or state and is also restricted by scarcity. The developable land is scarce because only 29 % of our planet is land and much of this 29 % is inaccessible or undevelopable due to topography. Other types of inputs are capital, materials, and also labor. Thus, the final cost of the construction is based on the cost of the building plots, wages of the workers, and cost of the materials (Égert and Mihaljek, 2007).

Other factors like building regulations, high permit fees, and administrative processes can directly affect profitability hence these institutional factors also affect the supply side. (Belke and Keil, 2017) Government monetary policies, like high-interest rates or taxation, may also have a profound impact on the housing market. The costs of all these factors have a negative effect on the amount of new construction, as with higher costs the profit gets smaller, thus there is a decrease in the motivation of investors and developers to provide new properties. Construction is a major element of the economy, accounting for 4.2 % of total GDP in the United States in 2021. At the beginning of 2022, over 7.5 million people worked in the construction industry.

3.3 Demand side

Based on conventional economic theory, the demand for properties can be defined as the number of units/number of houses demanded at various prices, where the lower amounts of units available, the higher the price. An important feature of the demand curve is price elasticity, which is determined by the availability of substitutes. Price elasticity is a notion that may be used to estimate the influence of changes in market prices or rents on demand, specifically the number of units demanded. It can also aid investors and developers in determining the revenue effect of price changes. Developers and investors favor inelastic demand because if prices or rents rise, revenues rise as well, as demand does not decrease enough to eliminate the gains from the price/rent increases.

Disposable income of households, which primarily consists of wages and salaries of the household, is considered to be the main determinant of the price. Data from BEA revealed the average real U.S. per capita income in 2021 was $55,711, representing a 6,03 % increase from the previous year. Wages and salaries are the two leading elements in the accumulation of wealth and savings of the household, thus increasing the possibility of purchasing a property or obtaining a mortgage loan to buy the property. Factors like the unemployment rate, number of vacancies, and the economic activity rate of the population can have a direct impact on disposable income whereas with a lower unemployment rate and greater economic activity the disposable income tends to get higher. On the other hand, when unemployment is greater consequently disposable income becomes lower and people can opt to rent properties instead of purchasing them, therefore the demand for real estate decreases, or the high rents can sway people into buying a property, thus, increasing demand for owner-occupied housing. Furthermore, because of "buy-to-let" arrangements,

the substitution effect between renting and owning can influence the demand for properties in both directions. (Belke and Keil, 2017) Therefore, variables like the unemployment rate, personal income per capita, and rental vacancy rates will be included in the analysis to further examine their effect on house prices.

3.3.1 Demographic factors

Furthermore, demand for housing is shaped by demographic variables such as population size or growth, population age structure, migration, the number of households, and overall changes in household size. Even features like marriage and divorce rates may have an impact on household formation and consequently housing demand. (Belke and Keil, 2017) These variables will be closely discussed in the practical part of this thesis.

The **population age structure** not only affects the demand for housing but also affects which type of dwellings are in demand. Finding new housing is most common for people aged from 20 to 35 years. These people are usually the ones who are increasing the demand for housing as in their early adulthood they want to move out of their parent´s house and live on their own or later they get married and look for a new house for their new family. (Asal, 2018) Similarly in research by Čadil (2009), a variable for the share of the population aged 20-39 was included, representing the sub-population that is most likely to increase demand for housing. Another variable that can come from the population age structure would be the economically active subpopulation which can be proxied by the share of the population aged 15-64, this group can also create demand for housing as they have jobs, therefore salaries to buy a property or they at least satisfy the conditions to apply for a mortgage loan.

Furthermore, **population growth** can be differentiated into net migration and natural increase. Both of these components were found to be significant by Hlaváček and Komárek. However, the natural increase had a smaller impact on demand as newborns are usually born into existing families and thus already existing households. On the other hand, net migration will increase the demand relatively immediately as new households are being created and therefore the prices will rise as well. As mentioned above, migration can also be an indicator of the economic strength of each state due to the reason that migration is often work-related.

Overall **changes in household size** affect demand as well. With the lower number of occupants per household, the demand for housing will increase and so will the prices,

making the housing even less affordable. Factors like *divorce and marriage rates* can play a role in determining the household size. Higher divorce rates mean a higher number of needed dwellings due to the divorce separating one household into two. Marriage rates were also tested but unlike divorce rates, they were not found to be significant, due to the possibility that weddings can cause either a merger of two households or create a need to buy a new dwelling for the family. In 2021, the average number of people per household in the United States was 2.51 and it is decreasing by the year. Moreover, the number of single-person households is still increasing which consequently increases the demand for dwellings. There were approximately 29 % of single-person households across the U.S. in 2019, with the highest percentage in the District of Columbia at 45 %, and the lowest in Utah, at 19 %. (Statista, 2022a) The marriage rate is constantly decreasing as the cohabitation of unmarried people is becoming more frequent. In 2020 only 5.1 people out of 1,000 were married, compared to the year 2016 when it was 7 people. (Cdc.gov, 2022)

From the demographic factors, variables share of the population aged 20-39 years and share of the population aged 15-64 years will be introduced and further described in Chapter 5 as well as marriage rates and population growth. Unfortunately, data on the average household´s size and divorce rates in each state were not available.

3.3.2 Financial market factors

Buying a residential property is often the largest purchase anyone can make in their life, and it often requires financing by a mortgage loan as well as some of their funds, thus the interconnection of the real estate market and financial factors is especially important. To ensure one makes an informed decision, it is important to understand the dynamics of real estate prices.

Long-term interest rates and mortgage rates can affect the affordability of a property, with growing mortgage credit the financing ability of a household increases and that can spur the demand for real estate. The higher the interest rates are, the worse the housing market will be. With higher interest rates, the debt of households increases, thus it dissuades homebuyers from borrowing, and as a result of this consumer demand is slowed. The conditions under which people can receive financing for their real estate are largely determined by their credit and loan availability. Mortgages are the most significant component of household debt. After the first quarter of 2022, the total household debt rose

to 15.84 trillion dollars out of which mortgage loans accounted for 11.18 trillion dollars (70.57 %). Since the end of 2019, debt increased by 1.7 trillion dollars.

The Federal Reserve (Fed) can influence the economy through monetary policies like setting the federal funds rate, which indirectly affects the mortgage rates. The federal funds rate is the rate at which U.S. financial institutions charge each other for loans of reserves deposited at the Fed. An increase in the federal funds rate makes it more expensive for banks to borrow money therefore they increase their interest rates to the customers and that puts upward pressure on mortgage interest rates as well. The Federal Reserve can decrease the mortgage interest rates by purchasing mortgage-backed securities, which they did during the 2007 Recession. During COVID-19 The Fed took similar actions, they purchased mortgage-backed securities and lowered the federal funds rate. Because of these actions in December of 2020 the 30-year fixed rate reached its historical low at 2.67 % and thus demand drastically increased and exceeded the national supply which increased the home price. (Weinstock, 2021) Based on this finding, the national 30-year fixed mortgage rate will be examined in the analysis.

4 Literature review

Research conducted by Capozza, Hendershott, Mack, and Mayer (2002) focused on the U.S single-family housing market. They explored the causes of differences in the dynamics of real house prices by using a large panel dataset with 62 metropolitan areas for the 17 years from 1979 to 1995 and estimating the mean reversion coefficient and serial correlation coefficient. Results revealed that house prices in different metropolitan areas vary not just because of the differences in local economies but also because they react differently to economic shocks which are dependent on factors like real construction costs, income, population growth, or area size. A higher serial correlation was found in metropolitan areas which have higher population growth, real construction costs, and income. On the other hand, higher construction costs are correlated with lower mean reversion. Based on empirical evidence, the overshooting of prices can appear in the areas with higher real construction costs. This can occur in the coastal cities like Los Angeles, New York, Boston, San Francisco, and San Diego where the mean reversion is low and serial correlation is high.

The next study also focuses on the United States, authors of this paper Holly, Pesaran, and Yamagata, (2010) use state-level data to conduct an empirical examination of changes in real housing prices in the U.S. The study looks at how much fundamentals, real per capita disposable income, influence real house prices at the state level and measures how quickly real house prices respond to macroeconomic and local disturbances. Additionally, the study carefully accounts for both cross-sectional dependence and heterogeneity. This allows them to find a cointegrating relationship between real house prices and real per capita disposable income and also to confirm the positive effect of population growth on house prices as well as a negative effect of net borrowing cost. In the analysis, the researchers included 49 states of the U.S. over a time period of 29 years. After the researchers properly accounted for both heterogeneity and cross-sectional dependence, an error correction model with a multifactor error structure and with a cointegrated relationship between real housing prices and real incomes was implemented. Their findings are consistent with the idea that real home prices have been increasing in line with economic fundamentals (real earnings) and that national housing bubbles are not particularly evident. However, some outlier states were discovered. Those states were: California, New York, Washington, Oregon, Rhode Island, Connecticut, and

Massachusetts. In those states, the logarithmic form of house price to income ratios was either unconnected to the U.S. average or even heading the other way.

In research carried out in 2009, researchers Clark and Coggin (2011) tried to answer the question of whether there was a U.S. house price bubble in the past through an econometric analysis using national and regional panel data. Specifically, by utilizing quarterly data from the first quarter of 1975 to the second quarter of 2005. They concentrated on the time series statistical relationship between the real U.S. and regional house prices and a variety of key economic indicators relevant to house prices. Firstly, they tested the stationarity of data and then the cointegration test revealed no cointegration relationship among the I(1) variables. Then they divided their data into two subperiods based on the dramatic rise in housing prices in the mid-1900s and they were then able to confirm that the increases in the housing prices from the first subperiod to the second are significant with the largest increases in the western areas of the U.S. and the smallest increases in the southern areas. Secondly, the researchers were able to confirm the existence of a house price bubble in the U.S. as a whole nation and in regional house prices through the use of Engle-Granger and Gregory-Hansen tests to identify a housing bubble.

In all of the above-mentioned studies, researchers focused on examining the determinants of house prices in the United States. However, in each study, the researchers chose disparate areas, metropolitan areas, states, and regions for their study. In the first two studies, disposable income and population growth were used and they were found to be important determinants of house prices. Hence, why these two variables were chosen to be the benchmark model in the analysis in Chapter 5.3. However, in the study by Holly, Pesaran, and Yamagata (2010), researchers used only 3 variables thus the significance of population growth could be influenced by the lack of other explanatory variables. In the same research, the authors used state-level data and found a cointegrated relationship between real income and real house prices. However, Clark and Coggin (2011) used regional data and found no cointegration among their variables. Due to this inconclusive result, the cointegration relationship among these two variables will be further tested in Chapter 5.2.3. Additionally, unemployment rate, 30-year fixed mortgage rate, homeownership rates, and debt-to-income ratio were used in the research by Clark and Coggin, (2011) and they did not find any cointegration relationship among their data, even though some of these variables were found to be I(1) for certain regions. This finding will be also tested in Chapters 5.2.1 and 5.2.3.

Similar research to the above-mentioned studies was written by Hlaváček and Komárek (2009), however, their paper was focused on regional determinants of real estate prices in the Czech Republic. Their paper is considered to be one of the first comprehensive studies on this topic in the Czech Republic. A decade later, this topic was further examined by Hlaváček and Kalabiška (2020). From the previous studies, this one differs mainly in the authors focus not only on defining the factors which are affecting the real estate prices but also focus on the dynamics of the various studied factors over time. In the first study, the intention was to identify the periods of overvaluation of real estate prices, which caused housing price bubbles in the market. For the analysis, authors used quarterly data from 1998 to 2009 for the Czech Republic as a whole and Prague, its capital city, and they managed to reveal two house price bubbles, first in 2002/2003 and the second one in 2007/2008. Lastly, for the analysis by regions Hlaváček and Komárek (2009) used fixed effect panel regression and simple OLS with first differenced non-stationary variables to establish the main drivers of apartment prices to be the demand factors. However, they used only annual data for 10 years, and thus, even authors advise using these results with caution due to the brief time span. In the study by Hlaváček and Kalabiška (2020), they examine the apartment prices based on quarterly data from 2000 to 2017 using the panel dynamic OLS estimator. Then an error correction model was employed to investigate the price adjustments and their speed in the short run and finally the existence of long-term equilibrium for the apartment prices was verified. In both of the above-mentioned studies, demand factors like unemployment rate and wages were found to be the most notable determinants. An additional finding by Hlaváček and Kalabiška (2020) was that building plot prices displayed unexpected negative effects in the low-income regions. Based on these two studies multiple variables which displayed significant effects in several regions were chosen for the analysis in this thesis and they will be presented in Chapter 5.1.

Another fairly recent study concerning real estate prices was written by Belke and Keil in 2017. In this study authors used annual data from 127 large German cities and regions in the years 1995 to 2009 which made a unique dataset, because the dataset was larger in the cross-sectional dimension than other datasets used in similar studies. The aim was to establish empirical facts about the determinants of real estate prices in Germany. They analyzed a panel dataset with a vast range of data covering the real estate market and other demographic, economic, and infrastructural variables. Two dependent variables were used in this analysis, firstly the apartment prices were defined as a price index for newly built

apartments and secondly the house prices which were substituted by purchasing prices of single-family houses. The authors used a two-way fixed effect panel model for the regression which identified several robust determinants. From the supply side, the number of newly constructed apartments and the number of transactions, which measures the activity of the real estate market, were found to be important determinants displaying a positive effect, the number of existing apartments was also found significant but with opposite sign. Demand factors that were found to be significant were age structure, with a positive effect, and regional infrastructure proxied by the number of hospitals per 1,000 inhabitants in each city. However, even though average annual disposable income per capita had a positive sign, it was not found significant in the estimations.

Last but not least, a paper written by Chien, Lee, and Cai (2014) who performed a panel cointegration analysis for macroeconomic determinants of the international housing market investigated the long-run and short-run dynamics. A quarterly dataset of 33 countries over the time period of 1980 to 2013 was used. After applying numerous stationarity tests, they tested for cointegration of I(1) variables using multiple tests but mainly Pedroni´s panel cointegration test. Based on the found cointegrating relationship between variables, they chose to use the panel dynamic OLS estimator which presented such results that with a 1 % increase in economic activity, construction costs, and long-term interest rates house prices would increase by 2.16 %, 0.22 %, and -0.04 % respectively. Another interesting finding implied that the increase in economic activity and construction costs has a higher impact on house prices in lower-income countries and that the highest coefficient of economic activity was found in the U.S.

Based on all of the above-mentioned studies the structure of this thesis was created, and the preliminary analysis was conducted. Appropriate variables, which displayed significant effect, and the data from the United States at the state level were available, were chosen for the analysis. However, even though fixed effect or panel dynamic ordinary least squares estimators are frequently used in the above-mentioned literature, a different model will be used for this thesis in the end because of the results from the stationarity and cointegration test. The model selection and the reasoning behind it will be further explained in Chapter 5.3.

5 Empirical analysis

In this section, the quarterly data for the analysis that covers 50 states of the United States in the 21 years between 2000-2020, will be presented. The method of collection, their purpose, and also the frequency in which they are publicly available. This chapter will be divided into sections based on which side of supply or demand the data influences.

5.1 Data selection

5.1.1 Dependent variable

Firstly, the dependent variable *"home value"* was taken from the housing data of Zillow Research. The Home Value Index (ZHVI) is a seasonally adjusted, smoothed estimate of typical home value and it reflects the appreciation and growth of home values across specific geographic areas and housing types. It represents the average price for properties in the 35th to 65th percentiles. This index allows more accurate tracking of home value changes over time compared to the S&P CoreLogic Case-Shiller home value index. The Case-Shiller home value index only includes homes that have sold, which is approximately 2 % of homes in a given year, and it measures changes in market value by comparing a home‘s most recent selling price to its last selling price. ZHVI is based on "zestimates". Zestimate uses home characteristics like square footage, the number of rooms and bathrooms, tax assessments and prior sales prices, and many other attributes. Afterward, through statistical and machine learning models they examine hundreds of data points for each home to estimate its current and historical market value. Reported zestimates are compared each month, allowing for monthly reports that are timelier than the Case-Shiller index. (Zillow Research, 2022)

5.1.2 Supply factors

As mentioned in Chapters 3.1 and 3.2 the supply of housing units consists of two parts, the existing units, and the newly constructed units. The same segmentation was used in the research by Hlaváček and Komárek (2009). Consequently, in this analysis, the variable housing units will be used as a proxy variable for the existing stock of houses with inelastic supply, and as a measure of construction of the newly built homes the variable building permits will be used.

The *housing units* variable comes from The Census Bureau's Population Estimates Program, and it is reported in the number of units per 1,000 inhabitants. Based on the study

by both Belke and Keil and by Hlaváček and Komárek this variable is expected to have a negative effect on the prices because a larger supply of housing lessens the relative demand pressure in the real estate market

Continuing with the *building permits* variable, which represents the number of issued building permits per 1,000 inhabitants of each state. This variable is derived from The Building Permits Survey from the United States Census Bureau. This survey provides monthly data on new privately-owned residential construction. However, the Monthly Building Permits Survey covers 8,400 permit-issuing places, out of approximately 20,000 permit-issuing places. These 8,400 permit-issuing places each issue at least 6 permits per year unlike the rest of the permit-issuing places and together they account for 99 % of the total units of new residential housing authorized each year. The reported data on building permits is seasonally adjusted. The expected sign of this variable can be either positive or negative. However, Belke and Keil found the number of newly built apartments per 1,000 inhabitants to be significant with a positive sign, probably due to the increased demand for housing causing new construction. (U.S. Census Bureau, 2022b)

In Chapter 3.2 another possible determinant was discussed. This determinant was the cost of construction and for the approximation of this variable, the *average valuation of construction per unit* reported in thousands of dollars will be used. This variable comes from the same survey with the same specifics as the building permits variable mentioned above. The valuation is the estimated value of the residential structure as indicated on the building permit. If there is no value indicated, the estimate is based on the average cost per unit for the same census region and the type of structure. The total data for permit valuation covers four sizes of residential buildings: single-family houses; two-unit buildings; three- and four-unit buildings; and buildings with five or five plus units. The reported data can be influenced by the fact that there is insufficient information to determine how closely the building permit valuation approximates the dollar amount of construction work involved. Because of the lack of data on the construction cost index at the state level, the average valuation from this survey will be used as a proxy. The construction costs data are available only at a national and regional level.

Subsequently, two variables that were not included in any previous research will be presented. Firstly, the *homeowner vacancy rates*, which are calculated by dividing the homeowner inventory that is vacant and for sale by the sum of owner-occupied units plus

units awaiting occupancy and the units which are vacant and for sale, and secondly, the *rental vacancy rates.* These are the portion of the rental inventory which is vacant for rent. Both of these variables should have a negative effect on house prices as with an increase in vacancy rates there will be a higher supply of housing units on the market either for sale or for rent, thus the prices should decline. Both of these above-mentioned vacancy rates are published quarterly by the Housing Vacancies and Homeownership survey.

Lastly, one more variable comes from this survey, which is the *homeownership rates* variable, which was included in the analysis by Clark and Coggin (2011). This variable is expected to have an opposite effect to the two beforementioned variables. The homeownership variable is calculated as a proportion of households that are occupied by the owners divided by the total number of occupied households.

5.1.3 Demand factors

As mentioned in Chapter 3.3, disposable income can be influenced by several factors which are all indirectly included in GDP. In the study made by Égert and Mihaljek, GDP per capita was used as a proxy for the disposable income of households, where they determined its significance. Another possible variable that can be chosen instead is the personal income per capita, however, this variable directly impacts the variable of GDP per capita, therefore, to avoid correlation, only one of these variables can be used in the analysis. Thus, the first variable on the demand side in this analysis will be *personal income per capita* which comes from the Bureau of economic analysis. The annual personal income consists of the income that U.S. residents get from paychecks, business ownership, employer-provided supplements like insurance, Social Security, and other government benefits, interests, and dividends, however, capital gains from changes in stock prices are not included. Then it is divided by the quarterly population estimates to obtain the per capita personal income, which could be a good indicator for assessing and comparing the economic well-being of state residents. (Bea.gov, 2022) This variable was chosen based on the empirical literature and it is expected to have a positive effect on house prices as mentioned in Chapter 4.

Subsequently, an additional variable that was discussed in Chapter 3.3 was the *unemployment rate* and its effect on disposable income. Thus, this can be a conceivably influential determinant of demand for housing. The unemployment rate displays the percentage of unemployed workers in the labor force, and it is an indicator of the local job

market and the economy´s strength. The monthly unemployment and labor force data for States, by place of residence, are produced by the Local Area Unemployment Statistics (LAUS) program. Based on the results of both Hlaváček and Komárek (2009) and Belke and Keil the negative effect of unemployment rates on prices was confirmed but it was found significant only by the first researchers.

Another variable related to income would be the debt-to-income ratio, which is calculated by The Fed as a ratio of median household debt and annual household income. It presents relevant information about the economic performance of different states. It was also included in the analysis by Clark and Coggin (2011), where they determined no cointegrated relationship among this variable.

The next variable chosen for this analysis offers an insight into the strength of a state's economy as well as consumer purchasing trends, that variable is the *personal consumption expenditures* (PCE) and it should approximate the consumer price index, which was studied in the research by Capozza, Hendershott, Mack and Mayer (2002). It covers all goods and services purchased by, or on behalf of, people living in each state. (Bea.gov, 2021)

Chapter 2.3 covered the population density and property taxes, both of which could have an impact on house prices in different states. The *population density* could play a role in determining the demand for housing, thus the expected effect would be positive, whereas with an increased demand the prices would increase as well. This variable was obtained by dividing the total quarterly population by the land area sizes of each state from the U.S. Census website. The quarterly population estimate dataset is available on the Bureau of Economic Analysis site and it is based on unpublished censuses. Due to the fact that property taxes are levied by each county or even neighborhood, there was no available state-level data. Thus, one of the best possible proxies available was the *tax burden*. The tax burden variable is calculated as a percentage of total income earned in a state which goes towards the payment of the aggregate amount of state and local taxes paid by residents. The aggregate amount of state and local taxes includes income and property tax, general sales tax, and many more categories and it is reported by the Census Bureau´s State and Local Government Finance division. Tax burden data comes from The Tax Foundation. (York and Walczak, 2022)

The following described variables were selected based on Chapter 3.3.1. In the research by Čadil (2009) the variable *share of the population aged 20-39* per 1,000 inhabitants was determined significant in its first-difference logarithmic form. For this analysis, the variable share of the population 20-39 was calculated according to the 5-year age groups from the intercensal population estimates. The same methodology applies to the *share of the population aged 15-64*, which is a proxy variable for economically active subpopulations as mentioned in the chapter about demographic factors. This variable was previously found significant with a positive effect in research by Hlaváček and Kalabiška.

Even though Hlaváček and Komárek (2009), and also Hlaváček and Kalabiška (2020) divided the population growth into two variables - net migration and natural population growth, and they were both found to be significant, in this analysis only the overall population growth will be used, considering the lack of state-level data on these two aforementioned variables. The *population growth* variable is derived from the quarterly percentage change in population estimates. These estimates are obtained by using the population base and adding natural increase and net migration, both international and domestic, and subtracting deaths. Dröes and van de Minne (2016) confirmed in their paper the positive relationship of population growth to property prices in the Netherlands.

The next possible determinant from demographic factors is divorce rates. Divorce rates were found significant but because of the scarcity of data, they will not be included in the analysis. Even though *marriage rates* were found insignificant in the literature about the Czech Republic, they were not tested in any relevant literature regarding the United States and that is the reason behind including it in this analysis and further testing.

Lastly, there will be one variable for which only data on the national level are available. The 30-year fixed mortgage rate was included in the analysis of U.S. housing by Clark and Coggin (2011) however, they included this variable as a ratio of the 30-year fixed mortgage rate and the adjustable rate. Furthermore, the authors studied only cointegrating relationships among their variables and they did not perform any regression analysis. The average national *30-year fixed mortgage rate* illustrates the cost of borrowing to finance a property and will be included to see the average quarterly mortgage rates. This quarterly data comes from FRED. No quarterly state-level data are available thus, this could be considered the ideal proxy. Because there are a lot of factors that have to be considered when obtaining a mortgage loan with a certain rate, it is impossible to get an accurate

mortgage rate. Factors like the applicant´s credit score, percentage of down payment, age of the applicant, the value of the property, loan type and term, maturity ,etc. play a role in determining the rate. However, the expected sign of this variable is negative based on research from the international housing market and the theoretical foundation.

5.2 Preliminary data analysis

For the econometric analysis, a panel dataset will be used. It covers a time period of 21 years from the first quarter of 2000 to the fourth quarter of 2020 and in combination with 49 states of the United States plus the District of Columbia in cross-section. The state of North Dakota was excluded due to the lack of data for the home value variable. This makes a total of 4200 observations for each variable. The complete list of variables and their summary statistics can be found in Table 5.1. The appendix also contains an integral summary table where additional information is added. This table includes variability over time and across states, as well as details on the states which take on the maximum or minimum values for each variable. While looking at Table A.1 one can see that the District of Columbia (D.C.) has a dominant position as a capital city and federal district.

Table 5.1. Summary statistics

Variable	Units	Mean	Median	St. Dev.
Home values	dollars	204849.55	178562.83	96892.91
Housing units	per 1000 inhabitants	436.60	436.91	32.73
Building permits	per 1000 inhabitants	0.36	0.30	0.24
Construction value per unit	thousands of dollars	165.96	163.00	43.45
Homeowner vacancy rates	percentage	1.94	1.90	0.80
Homeownership rates	percentage	68.22	69.35	6.32
Rental vacancy rates	percentage	8.68	8.30	3.18
Personal income per capita	dollars	41728.05	40378.50	10628.72
PCE per capita	dollars	34142.89	33600.12	7712.64
Unemployment rate	percentage	5.60	5.17	2.17
Tax burden rate	percentage	9.79	9.80	1.38
Debt-to-income ratio	ratio	1.61	1.56	0.41
Population density	citizens per km2	151.14	39.39	548.51
Share of population 20-39	per 1000 inhabitants	271.31	268.12	21.38
Share of population 15-64	per 1000 inhabitants	664.2	664.56	17.42
Population growth	qt percentage change	0.18	0.16	0.19
Marriage rates	percentage	7.93	6.90	5.90
S&P 500	percentage	0.62	1.20	0.02
30-year mortgage rate	percentage	5.00	4.88	1.32

5.2.1 Stationarity

Firstly, to choose a suitable model, the unit root test was performed to determine the non-stationary variables which could cause the occurrence of spurious correlation in the regression. Before using any formal test, each variable was plotted and after the visual inspection, the variable of marriage rate and share of the population aged 15-64 were found to have a downward trend and five other variables displayed an upward trend and thus probably being non-stationary. These variables are home values, personal income per capita, PCE, valuation of construction per unit, and housing units in some states. Because the graphs for all 50 states were unclear in their condensed forms, the charts were not added to the appendix section. Variables with a unit root might turn into trend stationary if the trend is eliminated or included in the regression model, thus, to remove the trend the first difference model or logarithmic transformation have to be used.

To test the stationarity of data multiple tests were performed. Firstly, the Levin-Lin-Chu (LLC) test, the main limitation of this test is the assumed cross-sectional independence in time series, it assumes the homogenous AR (1) coefficient for each unit in the panel data. However, it can be assumed and tested that the data are cross-sectionally dependent. LLC tests the null hypothesis of a unit root against an alternative hypothesis of homogeneous stationarity. As Levin et al. (2002) noted, their test performs better for moderate-size panel data, where N (number of individuals) lies between 10 and 250 and the number of observations T is between 25 and 250, with a condition that $N/T \rightarrow 0$. The Breitung unit root test also implements the autoregression between panels. However, as opposed to the Levin-Lin-Chu test, it assumes that both N and T converge to infinity. The next test performed was Im, Pesaran, and Shin unit root test, where a standardized t-bar test statistic is proposed, which is based on averaging the Augmented Dickey-Fuller statistics across the states. It enables the simultaneous use of stationary and non-stationary series. Additionally, this test considers group-specific heterogeneity and serial correlation in the dynamics and error variances. The limitation of this test is the standard normal distribution of t-bar statistics. Lastly, the Augmented Dickey-Fuller unit root test and Phillips-Perron test were performed. Both were available with two specifications, the Fisher type with Chi-squared distribution and the Choi version with the assumption of asymptotic normality for $N \rightarrow \infty$. Both tests are testing the null hypothesis of an individual unit root. Given the assumed cross-sectional dependence of the data, the second-generation unit root tests were also employed as they allow for the data to be cross-sectionally

dependent. Namely, the Cross-sectionally Augmented Im, Pesaran, and Shin unit root test (CIPS) and also the Cross-sectionally Augmented Dickey-Fuller test (CADF).

All tests yielded similar results (which can be found in Table A.2, Table A.3, and Table A.4) and based on which we can conclude that the previously mentioned variables are indeed non-stationary. These variables are home values, personal income per capita, the share of the population aged 15-64 years, and marriage rates. Additionally, the tax burden variable was found to be non-stationary as well. The second-generation tests together with the Breitung test derived similar results as the previously mentioned tests. However, they found additional variables to be non-stationary as well, namely the population density and share of population aged 20-39 years variables. As the results for these variables are fairly inconclusive, it was not possible to either reject or confirm the variables being stationary or non-stationary. Furthermore, logarithmic forms of price variables were also tested, this would allow us to interpret the results of regression as a percentage change. After testing the first differences of all the non-stationary variables, one can conclude that for all the variables their first differences are stationary except for home values and PCE (Table A.5 and Table A.6). Therefore, for the regression the first difference of the logarithmic form of home values will be used since that one proved to be stationary in first difference. In light of this, we can say that all of the above-mentioned variables in levels follow a unit root and thus are integrated of order one except the home values and PCE variables which are integrated of order two.

To avoid multicollinearity and to illustrate that the regressors do not have a perfect or exact linear relationship with one another in the short run, a correlation matrix was created. The results in Table A.7 show that the PCE variable is highly correlated with the personal income per capita variable, therefore the PCE variable will be excluded from any further analysis as personal income was found a significant determinant in the study by Case and Shiller (2004), Holly, Pesaran and Yamagata (2010) and others. When looking at the correlation matrixes of each region, as well as each division other variables were found to be problematic. In Midwest and West regions, the share of the population aged 20-39 years was found to be correlated with the number of units and in the Northeast and South, the same variable was found to be correlated with the homeownership rate, population density, and the share of the population aged 15-64 years. In the three regions, the variable of average valuation per unit also displayed correlating relationship with personal income per capita.

5.2.2 Granger causality

Due to the high number of variables that would enter the regression, the Granger causality test was performed to deduce the number of causal determinants in regression, similarly to the study by Cohen and Karpaviciute (2017), Čadil (2009) and Hlaváček and Kalabiška (2020).

Granger (1969) created a methodology for examining the causality between time series. It can be employed to determine whether X results in Y or vice versa. The fundamental tenet is that if past values of X are substantial predictors of the current Y, even though the past values of Y were included in the model, the X exerts a significant influence on Y. The simple model presented by Granger (1969) is as follows:

$$X_t = \sum_{j=1}^{m} a_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + \varepsilon_t$$

$$Y_t = \sum_{j=1}^{m} c_j X_{t-j} + \sum_{j=1}^{m} d_j Y_{t-j} + \eta_t$$

Where the null hypothesis states that lagged values of X do not explain the variation in Y or vice versa that Y_t does not Granger-cause X_t against the alternative hypothesis.

This test assumes that all variables used for this test are stationary. Since in the previous chapter, the variables share of the population aged 15-64 years, marriage rates, tax burden, the logarithmic form of home values, and income were found non-stationary, their first difference will be used in this test. The level form was used for the rest of the stationary variables and the variable which displayed mixed results were tested in both level and first differenced form. The number of lagged terms was chosen based on Liew (2004), who established that for a sample size larger than 960 observations the Hannan-Quinn criterion (HQC) has the highest probability of correctly estimating the true lag length of the AR process. After estimating the vector autoregression (VAR) model, the optimal number of lags was determined to be 5.

The results of the Granger causality test can be found in Table A.8 and it determined that the variables of construction value per unit, tax burden, and population density in either form are not causal determinants of home values. Another variable found not to be a casual determinant was the share of the population aged 20-39 years. Since this variable was found stationary by some unit root tests but non-stationary by the Breitung, CIPS, and CADF unit root tests, it was tested both in levels and first difference in the Granger

causality test. However, neither of them was found to Granger cause the home values. Given this result and also the results from the correlation matrix this variable will be further eliminated from the regression as well as population density.

The variable of construction value per unit was not found as a causal determinant of the home values. This could be explained by the fact that the values indicated on the building permits are often only estimates based on the values from the whole region and thus do not correspond with the construction values in each state. Another factor that could cause this variable not to be a causal determinant could be the fact that the average construction value per unit is calculated from the overall prices of all units, which include single-family homes which are counted as one unit as well as apartment buildings that include many units. Given that the construction value per single-family home is costlier than the construction value per one unit in an apartment building, in the state where apartment buildings are more often the construction values does not have to correspond with the construction values of home values variable.

5.2.3 Cointegration

In the previous chapters, multiple variables were confirmed to be integrated of order one, thus being non-stationary in levels. This issue can be removed by applying the first differences. However, treating the non-stationary variables with unit root in the multivariate model is not completely straightforward as there can be a linear combination of these integrated variables that can be stationary, thus they cannot move independently of each other for long and they are fluctuating around a long-run equilibrium. The cointegrating relationship among variables could be used to understand the long-run dynamics among those variables. Two tests were used to test the non-stationary variables for cointegration. The results can be found in Table 5.3.

As mentioned in Chapter 4, mixed results for the cointegrating relationship between house prices and income were found in the studies. Thus, this relationship will be tested by itself. From the test it can be confirmed that those two variables are in fact cointegrated as proposed by Holly, Pesaran, and Yamagata (2010), the results of the cointegration tests can be found in Table 5.2. Subsequently, the rest of the non-stationary variables will be added to the tests.

Table 5.2. Cointegration tests for home value and personal income

Kao test for cointegration	
Augmented Dickey-Fuller test	-4.0424 0***
Pedroni test for cointegration	
Modified PP test	6.6932 0***
Phillips-Perron test	5.9186 0***
Augmented Dickey-Fuller test	6.1589 0***

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level

Firstly, the Kao (1999) cointegration test, which follows the Engle-Granger two-step residual-based cointegration test, where the I (1) variables are used in the regression, but it specifies cross-section intercepts and homogenous coefficients on the first stage regressors. Afterward, the residuals are tested for stationarity by the unit root test, where I (0) stationary residuals imply cointegrated variables and I (1) residuals reveal no cointegration among variables. The number of lags included was determined by the Hannah-Quinn (1979) information criterion (HQC), which should outperform other criteria because of the size of the sample. The I (1) variables were separated into sets. First is all the variables that were declared I (1) by all the unit root tests and second, the extended set, where the variable of housing units was added. Kao test the null hypothesis of no cointegration against the alternative that all panels are cointegrated. Kao rejected the null hypothesis at 1 % significance for both sets declaring these above-mentioned variables cointegrated.

Secondly, the Pedroni test was performed, and it confirmed the presence of cointegration. Results from both of these above-mentioned tests can be found in Table 5.3. Pedroni (1999, 2004) test is also based on the Engle-Granger (1987) cointegration test with a null hypothesis of no cointegration in a heterogeneous panel (large T and N) with one or more non-stationary regressors.

Table 5.3. Cointegration tests

Cointegration tests	restricted set		full set	
	statistic	p-value	statistic	p-value
Kao test for cointegration				
Augmented Dickey-Fuller test	-3.1380	0.0009***	-3.3122	0.0005***
Pedroni test for cointegration				
Modified PP test	8.3654	0***	6.6742	0***
Phillips-Perron test	8.9731	0***	4.9939	0***
Augmented Dickey-Fuller test	10.3163	0***	6.4403	0***

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level

Last but not least, to find the number of cointegrating relations, the Johansen (1991, 1995) cointegration test was performed. This test reports two types of statistics, trace statistics, and the maximum eigenvalue statistics. Both of these statistics test the null hypothesis of no cointegration against the alternative of cointegration. The Johansen test was applied to the full set of non-stationary variables. Included variables were home values and personal income both in logarithmic form, housing units, the share of the population aged 15-64 years, tax burden, and marriage rates. This test rejected the null hypothesis of at most 1 cointegrating relationship among the full set of variables at the 1% significance level. Table 5.4 presents the findings.

Table 5.4. Johansen panel cointegration test

	trace test	prob.	max. eigen test	prob.
None	539.4673	0***	419.4419	0***
At most 1	216.9407	0***	168.2166	0***
At most 2	106.1699	0.3176	88.911	0.7786
At most 3	75.8224	0.9658	71.5242	0.986
At most 4	97.4039	0.5548	90.4802	0.6592
At most 5	78.1956	0.9477	78.1956	0.9477

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level

5.3 Model specifications

Given the results above and the confirmation of cointegration among the variables, the next step in the analysis will be the estimation of the long-run relations between the variables. The relevant empirical literature employs a variety of models. However, due to the nature of this data and the fact that a mixture of I (1) and I (0) variables were revealed by stationarity tests, no I (2) variable will be used in the analysis and 1 cointegrating relationship exists, the ARDL Pooled mean group (PMG) estimators developed by Pesaran et al. (1999) was chosen as the most appropriate technique for the dynamic panels, where it is possible to examine the speed of adjustment and the long-run level relationships. Through the addition of lags on short-run coefficients that may vary across cross-sections, the panel ARDL modeling approach corrects for endogeneity in the regressors. The lags were selected based on the Hannan-Quinn criterion (HQC), which has the highest probability of correctly estimated true lag length for a sample size larger than 960 observations.

The general ARDL model is specified as follows:

$$Y_{it} = \sum_{j=1}^{p} \delta_i Y_{i,t-j} + \sum_{j=0}^{q} \beta_i X_{i,t-j} + \mu_i + \varepsilon_{it}$$

Where i denotes cross-sectional unit, t represents time period, p and q denote optimal lags, where $p \geq 1$ and $q \geq 0$, X denotes independent variables, β and δ are the coefficients, γ is a constant, μ_i represents the fixed effect, and ε_{it} an error term. This model can be reparametrized as an error correction model:

$$\Delta Y_{it} = \theta_i (Y_{i,t-1} - \beta_i X_{i,t-1}) + \sum_{j=1}^{p} \delta_i \Delta Y_{i,t-j} + \sum_{j=0}^{q} \beta_i \Delta X_{i,t-j} + \mu_i + \varepsilon_{it}$$

Where β_i are the long-run parameters and θ_i is the error correction term (ECT). Significant ECT with the size of coefficient between -1 and 0 implies that there exists a long-run equilibrium and the size of ECT represents the speed of adjustment. However, the panel ARDL estimator places an emphasis on the homogeneity in the long-run estimators, thus only the long-run estimators will be presented, together with the error correction term to see whether there is an adjustment toward the equilibrium in the short run.

Due to the high number of variables that would enter the regression, the specific-to-general procedure as well as later the general-to-specific reduction procedure will be implemented. The first approach will start with a benchmark model of variables, and the extended model will be selected based on adding the remainder of variables one by one to the already tested benchmark and evaluating their significance, sign, and size of the coefficient. The benchmark variables will be personal income per capita in logarithmic form and population growth as they were found significant determinants by Holly, Pesaran, and Yamagata (2010) and Dröes and van de Minne (2016). The second approach will be applied to an unrestricted model and based on the statistical significance of variables; the extended model will be chosen. Similar procedure was implemented by Gallin (2003), Égert and Mihaljek (2007) as well as Hlaváček and Kalabiška (2019).

To begin with, the benchmark variables will be regressed on the variable home value in logarithmic form, and it will be estimated by a few standard panel regression estimators. Firstly, one-way (FE (1)) and two-way fixed effect (FE (2)) estimator, where one-way controls only for state-specific fixed effect, and the two-way also implements time fixed

effect than random effect estimator (RE) and lastly the Pooled OLS. All the reported results can be seen below in Table 5.5.

Table 5.5 Standard panel estimators

Home value	FE(1)	FE(2)	RE	Pooled OLS
Personal income per capita	0.7391 (0.0099) ***	1.5388 (0.0414) ***	0.7415 (0.0100) ***	1.1862 (0.0188) ***
Population growth	0.0954 (0.0140) ***	-0.0599 (0.0107) ***	0.0980 (0.0141) ***	0.5034 (0.0239) ***
Intercept	4.2811 (0.1064) ***	-4.1724 (0.4384) ***	4.2554 (0.1119) ***	-0.5350 (0.2006) ***
Adjusted R-squared	0.5779	0.2343	0.5779	0.4905
Rho	0.8585		0.7947	

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level
Values in parenthesis are the reported standard deviations.

Table 5.6 Panel dynamic estimators

Home value	PDOLS	FMOLS	ARDL
Personal income per capita	1.1231 (0.0014) ***	1.1347 (0.0008) ***	1.1480 (0.0018) ***
Population growth	1.1447 (0.0795) ***	0.6470 (0.1707) ***	0.6493 (0.0637) ***
Adjusted R-squared	0.4329	0.4784	

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level
Values in parenthesis are the reported standard deviations.

As reported in the findings, the two-way fixed effect estimator produces less credible coefficients and has a lower adjusted R-squared, thus a worse fit than the one-way FE. Furthermore, to test which estimator is superior, whether the Pooled OLS or RE, the Breusch and Pagan Lagrange multiplier test for random effects was performed. It tests the null hypothesis of Pooled OLS being preferred as there is not a significant difference across sections. The probability derived from this test was 0.0000, thus the null hypothesis can be rejected at a 1 % significance level in favor of the alternative hypothesis that the RE is the preferred model between these two. Continuing with testing of the standard panel regression estimators, the Hausman (1978) test for the comparison of fixed effect and random effect estimators was used. The null hypothesis of the Hausman test states that RE model fits best. However, from Table 5.7 can be seen that the p-value is 0, thus it can be

concluded that the preferred model would be one-way FE. The same conclusion can be drawn also from the rho-statistic, which displays the proportion of variation that is accounted for by individual-specific effects. A higher rho in FE suggests that the FE estimator is more accurate.

Table 5.7. Breusch and Pagan LM test and Hausman test

	statistic	p-value / prob>stats.
Breusch and Pagan LM test	94427.45	0***
Hausman Test	29.31	0.0000

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level

From Table 5.5 and Table 5.6, it can be seen that the benchmark variables are significant at the 1 % level, however, the coefficients vary considerably. By examining the results, it becomes evident how skewed the results can be when the incorrect model is chosen. Since in this analysis dynamic panel data is used, the standard panel data estimators are biased, a cointegrating relationship was found among the integrated variables and some variables are stationary at level meanwhile others only after taking the first difference, thus there is a mixture of I (1) and I (0) variables on the right side of the equation and therefore the Autoregressive-distributed lag (ARDL) model was employed.

When estimating the benchmark model by ARDL, the coefficient of log(income), which is the personal income per capita in the logarithmic form indicated that a 1 % increase in annual personal income would cause a 1.15 % increase in home values. The coefficient of population growth is smaller; however, this variable was used in levels and not in logarithmic form, thus needs to be multiplied by a hundred. Therefore, with an increase by 1 percentage point in quarterly population growth the home values would increase by approximately 65 %. However, given that this variable is calculated like a quarterly change in population, the median quarterly change is 0.016 which would imply that a quarterly increase of 0.016 causes approximately 1.04 % increase in home prices. Thus, the increased demand for housing may be the cause of this price increase.

Table 5.8. ARDL regression results – supply factors

Home value	(1)	(2)	(3)	(4)	(5)	(6)
Housing units		-0.0031 (0.0002) ***				
Building permits			1.5173 (0.1008) ***			
Homeowner vacancy rates				-0.2547 (0.0276) ***		
Homeownership rates					0.0112 (0.0023) ***	
Rental vacancy rates						-0.0361 (0.0061) ***
Personal income per capita	1.1480 (0.0018) ***	1.2611 (0.0099) ***	1.1220 (0.0024) ***	1.2036 (0.0065) ***	1.0757 (0.0150) ***	1.1856 (0.0057) ***
Unemployment rate						
Tax burden						
Debt-to-income ratio						
Share of population 15-64						
Population growth	0.6493 (0.0637) ***	0.7579 (0.05121) ***	0.0556 (0.0808)	0.5927 (0.0969) ***	0.3762 (0.0560) ***	0.4259 (0.0921) ***
Marriage rates						
30-year fixed mortgage rate						
ECT	-0.0066 ***	-0.0095 ***	-0.0073 ***	-0.0061 ***	-0.0078 ***	-0.0062 ***

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level
Values in parenthesis are the reported standard deviations.

Table 5.9. ARDL regression results – demand factors

Home value	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Housing units							-0.0039 (0.0008) ***
Building permits							2.9467 (0.4427) ***
Homeowner vacancy rates							-0.2203 (0.0343) ***
Homeownership rates							0.0191 (0.0046) ***
Rental vacancy rates							-0.0331 (0.0080) ***
Personal income per capita	1.1243 (0.0033) ***	1.0186 (0.0219) ***	1.3019 (0.0270) ***	1.3557 (0.0497) ***	1.1761 (0.0036) ***	0.9385 (0.0258) ***	1.5450 (0.1040) ***
Unemployment rate	-0.0349 (0.0053) ***						-0.0616 (0.0117) ***
Tax burden		-0.1367 (0.0151) ***					-0.0348 (0.0223)
Debt-to-income ratio			-0.8064 (0.1488) ***				-0.0452 (0.0783)
Share of population 15-64				0.0029 (0.0008) ***			-0.0015 (0.0012)
Population growth	0.9053 (0.0400) ***	0.6805 (0.0746) ***	0.8369 (0.1354) ***	0.1417 (0.0797) *	0.2078 (0.1134) *	0.0059 (0.0025) **	1.7668 (0.3962) ***
Marriage rates					-0.0022 (0.0034)		
30-year fixed mortgage rate						-0.0557 (0.0121) ***	-0.0317 (0.0116) ***
ECT	-0.0093 ***	-0.0067 ***	-0.0039 ***	-0.0078 ***	-0.0059 ***	-0.0055 ***	-0.0063 ***

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level
Values in parenthesis are the reported standard deviations. (13) ARDL (2,1,….,1)

5.4 Results

As mentioned above when deciding what variables to use in the extended model, the testing started with model (1) (Table 5.8), the benchmark of two variables that were previously tested in studies about the United States. These variables were found significant with the correct expected sign. The analysis continued with testing the variables from the supply side, corresponding with models (2) to (6). These models revealed that all of the variables from the supply side are significant, where housing units, homeowner vacancy rates, and rental vacancy rates have a negative effect on home values and building permits have a positive effect on home values.

With an increase in housing units, there would be a 0.31 % decrease in home values. A decrease in homeowner vacancy rates and rental vacancy rates means that there are fewer vacant homes or fewer homes for sale or rent on the market, thus the price increase is caused by insufficient supply and raised demand. The homeownership rate has a positive effect on home values, which implies that with a higher number of owned homes, there is a lesser availability on the market which drives the prices upwards. The size of the effects and possible reasoning behind them will be explained further in Chapter 5.5.

A higher number of building permits obtained could mean that there was high demand for housing which incentivized developers to build more to increase the supply. However, due to the construction lag, it takes time to increase the supply. Based on the data from the U.S. Census Bureau (2022a), it takes approximately 9 months on average to build a house with one unit. Thus, even though building permits will increase the supply, it will happen eventually, and in the meantime, there is still pressure on the prices from the demand side and subsequently, the prices increase. The coefficient from the regression illustrates that with an increase by one, the prices would increase by 150 %, however, one must take a look at the units of this variable. For example, in Alabama in the first quarter of 2000, the number of building permits per 1,000 inhabitants was 0.3576. The number of building permits obtained was 1,591 and there were nearly 4.5 million people living in the State. Therefore, to have a one unit increase in building permits per 1,000 inhabitants it would mean that more than 6,000 building permits would have to be issued in the next quarter, which means 3.8 times the amount compared to the last quarter. Looking at the data from the dataset the median quarterly change is 0.006 which would indicate a 0.91 % increase in prices.

Models (7) through (12) in Table 5.9 examine the effects of demand factors. Firstly, in model (7) the variable of the unemployment rate was tested. The variable displayed the expected sign and was found to be significant with a negative effect on home values of 3.5 %, which would support the theory that with a decrease in disposable income of a household, the prices decline. Comparable results were found by Hlaváček and Komárek (2009) as well as Hlaváček and Kalabiška (2020) in the studies about regional determinants of real estate prices in the Czech Republic.

Afterward, in model (8), the tax burden variable was added and was found significant. Being that tax burden is calculated like a percentage of state income, it would mean that in order to increase the tax burden rate either an income would have to decrease, or the taxes would have to increase considerably. This goes hand in hand with the already proven positive effect of income on home values from studies mentioned in Chapter 4.

The analysis continued with the debt-to-income ratio, where the effect demonstrated the following, an increase in the debt-to-income ratio causes a decrease in home value of 80,6 %, which means that if the median debt of households increases or the annual household's income decreases the home values would decrease as well. However, the average quarterly change in debt-to-income ratio is 0.025, which would indicate that with a quarterly increase of this size, there would be a 2.02 % decrease in home values.

Next two variables, the share of the population aged 15-64 and population growth were found significant with the correct sign, which displays a positive effect on home values, explained by the increased demand for housing. However, the share of the population variable has a relatively low coefficient. House prices tend to get higher with increased demand, although it takes some time to satisfy the supply side as the supply of new properties adjusts to demand slowly. This leaves the equilibrium neither stable nor efficient in the short run and it could be one of the reasons a home price bubble may occur, due to the initial lack of supply which will later reverse into excess supply as the demand falls.

The last significant variable which will be included in the extended model based on the specific-to-general approach is the 30-year fixed mortgage rate which shows a negative effect of 3.17 %. The reason behind the decrease in home values could be that fewer individuals choose to borrow money to buy a home as mortgage rates rise; as a result, demand falls, and values follow.

Overall, from the specific-to-general approach, there were 12 variables included in the extended model - housing units, building permits, homeowner vacancy rates, homeownership rates, rental vacancy rates, personal income per capita in logarithmic form, unemployment, tax burden, debt-to-income ratio, the share of the population aged 15-64 years, population growth and lastly 30-year fixed mortgage rate. However, variables of tax burden, debt-to-income ratio, and share of the population aged 15-64 years lost their significance in the extended model because other variables in the model are probably of higher importance. All the other variables kept their significance at the 1 % level. The sign of the coefficient remained the same for all the variables, although due to the complexity of the regression model, the size of the coefficients is likely to change.

Subsequently, the analysis proceeded with the opposite approach (general-to-specific, Table 5.10). Starting with all the variables in the model, and then removing insignificant variables. The examination began with model (14) where three variables were found to be insignificant - tax burden, debt-to-income ratio, and share of the population aged 15-64 years. When removing the variables one by one, the coefficients of all the other variables did not considerably alter. Due to this approach, the variable of marriage rates was found significant and thus included in the final model, model (17). Surprisingly, the marriage rates turned out to be significant and with a negative effect on home values, whereas with a 1 % increase in marriage rates there would be a 1.31 % decrease in home values. This effect would imply that perhaps after marriage, the couple moves in together from two separate homes into one, thus increasing the supply on the housing market rather than decreasing it by purchasing another home together.

The error correction term revealed its significance and negative sign for all the models implying that there is indeed an adjustment towards the equilibrium. The ECT ranged from -0.0039 to -0.0095 which would demonstrate that the average speed of adjustment is 0.65 % per quarter, making it a 2.6 % annual adjustment.

Table 5.10. ARDL regression result - general-to-specific approach

Home value	(14)	(15)	(16)	(17)
Housing units	-0.0035 (0.0007) ***	-0.0032 (0.0006) ***	-0.0025 (0.0006) ***	-0.0041 (0.0007) ***
Building permits	2.5585 (0.3528) ***	2.3031 (0.2955) ***	2.2054 (0.2426) ***	2.4781 (0.2684) ***
Homeowner vacancy rates	-0.1838 (0.0282) ***	-0.1706 (0.0251) ***	-0.1832 (0.0248) ***	-0.1864 (0.0301) ***
Homeownership rates	0.0162 (0.0038) ***	0.0147 (0.0034) ***	0.0201 (0.0035) ***	0.0100 (0.0038) ***
Rental vacancy rates	-0.0331 (0.0080) ***	-0.0272 (0.0062) ***	-0.0279 (0.0059) ***	-0.0279 (0.0068) ***
Personal income per capita	1.4540 (0.0804) ***	1.3750 (0.0593) ***	1.4533 (0.0593) ***	1.3499 (0.0254) ***
Unemployment rate	-0.0501 (0.0093) ***	-0.0432 (0.0078) ***	-0.0331 (0.0065) ***	-0.0711 (0.0083) ***
Tax burden	-0.0130 (0.0182)			
Debt-to-income ratio	-0.0088 (0.0648)	-0.0330 (0.0568)		
Share of population 15-64	-0.0004 (0.0010)	-0.0002 (0.0009)	-0.0011 (0.0010)	
Population growth	1.4698 (0.3270) ***	1.3008 (0.2807) ***	0.8599 (0.2530) ***	1.1888 (0.2928) ***
Marriage rates	-0.0139 (0.0052) ***	-0.0128 (0.0045) ***	-0.0156 (0.0047) **	-0.0131 (0.0042) ***
30-year fixed mortgage rate	-0.0303 (0.0101) ***	-0.0284 (0.0092) ***	-0.0197 (0.0090) **	-0.0312 (0.0102) ***
ECT	-0.0073 ***	-0.0079 ***	-0.0075 ***	-0.0062 ***

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level
Values in parenthesis are the reported standard deviations. (17) ARDL (2,1,….,1)

5.5 Analysis by divisions

Firstly, due to the fact that datasets used for this chapter changed considerably, the stationarity of data for each division was further examined. An overview of U.S. divisions and regions can be seen in Table A.9 and their descriptive statistics can be found in Table A.14.,Table A.15.,Table A.16.,Table A.17. andTable A.18. Two unit root tests were performed, the Im-Pesaran-Shin (IPS) and the cross-sectional dependent Im-Pesaran-Shin (CIPS) unit root test. Results from the unit root tests are presented below and they can be found in Table A.10, Table A.11, and Table A.12.

For all the divisions, variables personal income per capita, tax burden, share of the population aged 15-64 years, marriage rates, and for 8 of the divisions home values were found to be non-stationary by the Im-Pesaran-Shin test. Housing units as well as home values in levels and tax burden were found to be non-stationary for all the divisions by the cross-sectionally dependent Im-Pesaran-Shin unit root test. Building permits, homeowner vacancy rates, homeownership rates, and rental vacancy rates were found stationary for all divisions by both tests and population growth was stationary for all divisions but one, the West South Central. For the rest of the variables, the results were rather mixed. For divisions (4) to (8), the CIPS test revealed that unemployment is also non-stationary

After checking for correlation with the use of correlation matrixes, variables PCE, population density, construction value per unit, and share of the population aged 20-39 years were found to be correlated with other explanatory variables in each division, therefore they were excluded from further analysis.

Subsequently, a cointegration test was employed for each division based on their results from stationarity tests. For all the divisions, the null hypothesis of no cointegration was rejected at least at a 5 % significance level, thus confirming the existence of a cointegrating relationship among variables. The results of the cointegration test by divisions can be found in Table A.13.

In the next section, the analysis continued with the extended model (17) with 10 variables, and it was applied to the data from each division. Table 5.11 and Table 5.12 represents the regression results for separate divisions. Models (1) to (3), represent divisions East North Central, East South Central, and Mountain division. Following, models (5) and (6) present the results from the New England and Pacific division,

respectively. Model (7) represents the South Atlantic division, which consists of 9 states, however, one of them is considered to be the District of Columbia, which was based on the summary table identified as a potential outlier due to its unique characteristics, thus for the purpose of this analysis D.C. was excluded from the South Atlantic division. Lastly, model (8) and (9) represent West North Central and West South Central.

The District of Columbia (D.C.) was found to have the lowest number of building permits per 1,000 inhabitants, the lowest homeownership rates, and construction value per unit. In D.C, there is a higher number of apartments than single-family houses, which are usually more costly to build than apartment units. According to the D.C. Policy Center housing database from 2017, there were approximately 226 thousand apartments, condominiums, and coops and only 93.5 thousand single-family homes. (edscape.dc.gov, 2018) Continuing with the lowest values, the debt-to-income ratio is the lowest in D.C. as well. This could be explained by the low homeownership rates which are at 40 % in the District of Columbia. That is 30 % less compared to the mean in other states. This indicates that residents of D.C. prefer to rent their homes and therefore they do not accrue any debt from obtaining mortgage loans. Moreover, in D.C. the personal income per capita reaches the maximum value in the entire United States. Lastly, the population density is the highest with more than 4,500 people per square km. Based on all the above-mentioned reasons D.C. was excluded as an outlier from the South Atlantic division.

Table 5.11. ARDL regression results by division - divisions (1) and (5)

Home value	(1)	(2)	(3)	(4)	(5)
Housing units	-0.0023 (0.0023)	-0.0028 (0.0019)	-0.0018 (0.006) ***	-0.0046 (0.0019) **	0.0011 (0.0011)
Building permits	0.7817 (0.3457) **	0.8564 (0.1687) ***	0.7614 (0.1454) ***	0.4521 (0.1065) **	0.8697 (0.3449) **
Homeowner vacancy rates	-0.0423 (0.0320)	-0.0623 (0.0197) ***	-0.0592 (0.0235) **	0.0443 (0.0510)	-0.1228 (0.0536) **
Homeownership rates	-0.0044 (0.0087)	0.0111 (0.0051) **	-0.0074 (0.0045)	0.0071 (0.0056)	0.0163 (0.0074) **
Rental vacancy rates	-0.0026 (0.0087)	-0.0014 (0.0047)	-0.0018 (0.0068)	-0.0367 (0.01834) **	-0.0233 (0.0210)
Personal income per capita	1.2703 (0.1007) ***	1.1645 (0.0610) ***	1.2702 (0.0371) ***	1.1901 (0.0555) ***	1.1984 (0.0389) ***
Unemployment rate	0.0167 (0.0105)	-0.0039 (0.0067)	0.0063 (0.0085)	-0.1105 (0.0197) ***	-0.0506 (0.0134) ***
Population growth	1.2715 (0.6192) **	-0.0858 (0.2051)	0.1411 (0.1667)	2.0609 (0.5800) ***	1.9805 (0.6317) ***
Marriage rates	-0.0095 (0.0043) **	-0.0034 (0.0082)	-0.0112 (0.0024) ***	0.0316 (0.0343)	-0.0285 (0.0364)
30-year mortgage rate	-0.0060 (0.0126)	0.0168 (0.0109)	0.0082 (0.0118)	-0.0027 (0.0183)	0.0216 (0.0186)
ECT	-0.0178 ***	-0.0180 ***	-0.0198 ***	-0.0116 ***	-0.0157 ***

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level
Values in parenthesis are the reported standard deviations.

Table 5.12. ARDL regression results by division - divisions (6) to (9)

Home value	(6)	(7)	(8)	(9)
Housing units	-0.0020 (0.0027)	-0.0081 (0.0010) ***	0.0012 (0.0024)	-0.0028 (0.0020)
Building permits	2.2378 (0.5243) ***	1.2660 (0.2772) ***	1.0642 (0.2959) ***	0.7513 (0.2468) ***
Homeowner vacancy rates	-0.1712 (0.0709) **	-0.0093 (0.0264)	-0.0655 (0.0282) **	-0.0345 (0.0200) *
Homeownership rates	-0.0318 (0.0200)	0.0055 (0.0048)	0.0196 (0.0062) ***	0.0074 (0.0066)
Rental vacancy rates	0.0166 (0.0191)	-0.0034 (0.0074)	-0.0052 (0.0086)	0.0037 (0.0069)
Personal income per capita	1.3411 (0.0367) ***	1.3734 (0.0339) ***	0.9541 (0.1011) ***	1.1444 (0.0786) ***
Unemployment rate	-0.0581 (0.0171) ***	-0.0292 (0.0096) ***	0.0017 (0.0128)	0.0145 (0.0107)
Population growth	0.1421 (0.4668)	-0.2573 (0.3633)	0.3059 (0.4556)	0.4047 (0.1852) **
Marriage rates	0.0293 (0.0090) ***	0.0159 (0.0171)	-0.0337 (0.0255)	0.0157 (0.0092) **
30-year mortgage rate	-0.0315 (0.0184) *	-0.0282 (0.0114) **	-0.0049 (0.0122)	-0.0068 (0.0114)
ECT	-0.0147 ***	-0.0119 ***	-0.0155 ***	-0.0096 ***

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level
Values in parenthesis are the reported standard deviations.

For the majority of the significant variables, the sign remained the same; however, the size of the coefficients changed substantially across divisions. For all the divisions there was a considerably high effect of building permits on the home values which would imply a high pressure on the prices from the supply side. An increase in the number of building permits per 1,000 inhabitants is connected on average with a 107 % increase in home values. The highest effect of building permits was found in the Pacific division (224 %) and the second highest (127 %) in the Mid-Atlantic division. Although the influence appears to be impactful, one must keep in mind the variable's units. The average quarterly change in the Pacific region is 0.004 which would indicate only a 0.9 % increase in home values.

Housing units turned out significant for the Mountain division, the Mid-Atlantic division, and the South Atlantic division with a negative and relatively small effect on home values. A negative sign was anticipated since a greater supply in the housing market lessens the pressure from the demand side. Homeownership rates remained significant and with the correct positive sign for three out of eight models, for the divisions East South Central, New England, and West North Central, where with a 1 % increase in homeownership rates there would be a 1 to 2 % increase in home values. The highest coefficient was reported in West North Central which had the highest homeownership rates in 2020 at 70 %. This could insinuate an increased demand for ownership of homes in this division thus, the leading effect of the aforementioned variable on home values in this division.

Homeowner vacancy rates were found to be significant for the majority of the divisions except for East North Central, South Atlantic, and Mid-Atlantic, and the effect ranged from 3.45 % to 17.12 % in West South Central and Pacific divisions, respectively. An increase in homeowner vacancy rates would imply a higher supply of homes on the market, therefore the decline in values. The average homeowner vacancy rate in the Pacific division in quarter four of 2020 was 0.76 % thus an increase by 1 % would mean a 2.32 times higher vacancy rate than the previous quarter. This would create an excess supply based on which the values would decrease. However, the median quarterly change is 0.1 %, which would indicate a 1.7 % decrease in home values. For the East North Central, the average rate is 1.35 % thus a 1 % increase would mean a 1.74 times higher vacancy rate than the previous quarter.

Rental vacancy rates did not maintain their significance for any of the divisions except one, the Mid-Atlantic, where rental vacancy rates had a negative effect on the home values of a magnitude of 3.67 %. However, for the rest of the divisions, the results indicated no significant effect of vacant rental properties on the home values. The reason behind it could be the fact that the ownership of a home is still considered to be part of the American dream, thus people prefer to purchase a home rather than rent it. Data from the 2019 American Housing Survey confirmed this notion, given that the national homeownership rate was 64 % and the rental rate was almost twice as low at 36 %. (U.S. Census Bureau, 2019). However, the rental rate in the Mid-Atlantic division and the Pacific division is 4 % and 7 % above the national average, respectively. This indicates that renting in these divisions is more frequent, thus it could potentially affect the home values. However, in the Pacific division, the rental vacancy rate was not found significant, due to the fact that other variables were probably of higher importance.

On the other hand, personal income per capita yielded significant coefficients for all the divisions, with the highest values in the Pacific and South Atlantic divisions; approximately a 1.35 % increase in home values with a 1 % increase in personal income per capita, and the lowest value in West North Central; a 0.95 % increase with a 1 % increase in personal income per capita.

The unemployment rate and population growth were found to be significant in only four divisions. The unemployment rate and population growth proved to be significant for the New England division, where a higher unemployment rate by 1 % would cause a 5 % decrease in home values and a 1 % increase in quarterly population growth would cause a 198 % increase in home values. However, the median quarterly change in population growth for this division is approximately 0.015, which would imply a 2.97 % increase in prices. A similar effect of population growth was revealed in the Mid-Atlantic where the regression also determined the highest effect of unemployment rate where with an increase in the unemployment rate by 1 % the prices would decline by 11 %.

An interesting finding was that marriage rates have opposite effects on home values in different divisions. They were found significant on at least a 5 % significance level in four divisions. In the Mountain division and East North Central, a 1 percent increase in marriage rates would cause a 1.12 % and 0.95 % decrease in home values. These two divisions are in the mid-range of home values, thus probably the most frequent decision after a wedding

is to move together into the home in which one of the persons from the couple already lived. As a result, the supply in the housing market increases hence the decrease in home values. Meanwhile, in the Pacific division and West South Central, the effect is positive, implying that an increase in marriage rates causes a 2.9 % and a 1.6 % increase in home values, respectively. Given that the Pacific division is the most expensive division when it comes to real estate, a single person might incline towards living in a small apartment, however, after a wedding, the couple wants to purchase a home together thus increasing the demand for housing. On the contrary, the West South Central division is the second cheapest after East South Central, therefore the explanation could be that even though each person lives in their own home before the wedding, they want to jointly purchase a new home after the wedding which increases the demand.

The national 30-year fixed mortgage rate displayed a significant effect only in two divisions although for the Pacific division it was only significant at a 10 % significance level. If the 30-year fixed mortgage rate rises by 1 % the home values would decrease by 3 %, due to the fact that the cost of borrowing to finance a property would increase, and thus the demand for housing would decrease and therefore the values would decrease as well.

Last but not least, the error correction term was found significant and in the correct range of (-1,0). For most of the models, the coefficient of ECT fluctuated at around -0.015. The highest speed of adjustment was reported by the Mountain division at a 1.98 % adjustment toward the equilibrium each quarter, nearly 8 % adjustment annually and the lowest value was reported by West South Central at barely 1 % speed of adjustment every quarter.

6 Conclusion

This thesis aimed to examine the effect of macroeconomic and housing indicators on real estate prices in the United States. The robust evaluation of long-term macroeconomic effects was made possible with the use of the panel dataset described in Chapter 5. After conducting multiple first and second-generation stationarity tests, the dependent variable was found to be non-stationary as well as a number of the independent variables. However, many variables from the supply factors turned out to be stationary. Thus, because of the mixture of different orders of integration and also due to the fact that Pedroni and Kao cointegration tests revealed cointegration among variables, and the Johansen cointegration test confirmed only a single cointegrated relationship among variables, the Pooled mean group ARDL model was employed. The ARDL model was later reparametrized as an error correction model and estimated in order to determine the speed of adjustment towards the long-term equilibrium. Other standard panel data estimators like Pooled Ordinary Least Squares, Fixed effect, and Random effect model as well as dynamic panel data estimators - Fully Modified Ordinary Least Squares and Dynamic Ordinary Least Squares, were also tested on the benchmark model. Due to the high number of variables the analysis started by testing the benchmark model with only two variables, which were found significant in the study by Holly, Pesaran, and Yamagata (2010). After testing the variables one by one and further employing the opposite approach of testing, the extended model was specified.

Finally, this extended model was used to determine the size of the impact of numerous factors on property prices in each division. For all the divisions, personal income per capita and building permits were found to be significant determinants. From the supply factors, homeowner vacancy rates were found to be significant for six out of nine divisions, which simultaneously with the effect of building permits implies that the supply of housing units is a strong determinant of real estate prices. In addition to personal income, it was discovered that demand factors such as unemployment rates, population growth? and even marriage rates were significant determinants, each being significant in four divisions.

This research could be possibly further extended by testing metropolitan areas, considering that in the United States nearly 83 % of the population lives in cities or urban areas, thus the data on metro area levels could be more cogent. However, the data available on this level is fairly limited. Furthermore, the addition of other variables into the regression could enhance the study. The interconnection of the housing market and

financial market could be examined through the inclusion of, for example, interest rates, other types of mortgage rates as well as exchange rates to include the international competitiveness factor. Another interesting topic could be to analyze the recent years, whether a new housing bubble occurred during the Covid-19 pandemic and its effect on house prices.

Bibliography

Adams, Z. and Füss, R. (2010). Macroeconomic determinants of international housing markets. *Journal of Housing Economics*, 19(1), pp.38–50. doi:10.1016/j.jhe.2009.10.005.

Asal, M. (2018). Long-run drivers and short-term dynamics of Swedish real house prices. *International Journal of Housing Markets and Analysis*, 11(1), pp.45–72. doi:10.1108/ijhma-08-2017-0070.

Balemi, N., Füss, R. and Weigand, A. (2021). COVID-19's Impact on Real Estate markets: Review and Outlook. *Financial Markets and Portfolio Management*, 35(4). doi:10.1007/s11408-021-00384-6.

Balint, N. (2018). *Goodbye LA, Hello Phoenix: Affordable Housing Makes Second-Tier Metros Top Choices for Moving*. [online] RentCafe. Available at: https://www.rentcafe.com/blog/rental-market/real-estate-news/top-10-counties-where-people-are-moving-offer-half-price-housing/.

Barbieri, L. (2005). Panel Unit Root Tests: A Review. [online] Available at: https://www.researchgate.net/publication/252756953_Panel_Unit_Root_Tests_A_Revie

Bea.gov. (2021). *Consumer Spending by State | U.S. Bureau of Economic Analysis (BEA)*. [online] Available at: https://www.bea.gov/data/consumer-spending/state.

Bea.gov. (2022). *Personal Income by State | U.S. Bureau of Economic Analysis (BEA)*. [online] Available at: https://www.bea.gov/data/income-saving/personal-income-by-state.

Belke, A. and Keil, J. (2017). Fundamental Determinants of Real Estate Prices: A Panel Study of German Regions. *International Advances in Economic Research*, 24(1), pp.25–45. doi:10.1007/s11294-018-9671-2.

Belke, A.H. and Wiedmann, M. (2005). Boom or Bubble in the US Real Estate Market? *Intereconomics*, 40(5), pp.273–284. doi:10.1007/s10272-005-0157-0.

Bureau of Labor Statistics (2021). *CONSUMER EXPENDITURES - 2020*. [online] Available at: https://www.bls.gov/news.release/pdf/cesan.pdf.

Čadil, J. (2009). Housing price bubble analysis - case of the Czech republic. *Prague Economic Papers*, 18(1), pp.38–47. doi:10.18267/j.pep.340.

Capozza, D., Hendershott, P., Mack, C. and Mayer, C. (2002). Determinants of Real House Price Dynamics. *NBER Working Paper No. w9262*. doi:10.3386/w9262.

Case, K.E. and Shiller, R.J. (2003). Is There a Bubble in the Housing Market? *Brookings Papers on Economic Activity*, 2003(2), pp.299–362. doi:10.1353/eca.2004.0004.

Cdc.gov. (2022). *NVSS - Marriages and Divorces*. [online] Available at: https://www.cdc.gov/nchs/nvss/marriage-divorce.htm.

Census Regions and Divisions of the United States. (2018). [online] Available at: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf.

Chen, J. (2020). *What Are Real Estate Market Tiers?* [online] Investopedia. Available at: https://www.investopedia.com/terms/r/real-estate-tier-classifications-tier-1-tier-2-and-tier-3.asp.

Chien, M.-S., Lee, C.-C. and Cai, S.-J. (2014). A Panel Cointegration Analysis for Macroeconomic Determinants of International Housing Market. *International Journal of Economics and Management Engineering*, 8(7).

Chirwa, Themba G & Odhiambo, Nicholas M., 2018. The determinants of public debt in the Euro area: A panel ARDL approach. *Working Papers 23509, University of South Africa, Department of Economics.*

Clark, S.P. and Coggin, T.D. (2011). Was there a U.S. house price bubble? An econometric analysis using national and regional panel data. *The Quarterly Review of Economics and Finance*, 51(2), pp.189–200. doi:10.1016/j.qref.2010.12.001.

Dröes, M. and van de Minne, A. (2016). Do the Determinants of House Prices Change over Time? Evidence from 200 Years of Transactions Data. *25th Annual European Real Estate Society Conference*. doi:10.15396/eres2016_227.

edscape.dc.gov. (2018). *Types of Housing* [online] Available at: https://edscape.dc.gov/page/neighborhood-factors-types-housing.

Égert, B. and Mihaljek, D. (2007). Determinants of House Prices in Central and Eastern Europe. *Comparative Economic Studies*, 49(3), pp.367–388. doi:10.1057/palgrave.ces.8100221.

El Nasser, H. (2018). *What is Rural America?* [online] The United States Census Bureau. Available at: https://www.census.gov/library/stories/2017/08/rural-america.html.

Eviews.com. (2020). *Estimating ARDL Models In EViews*. [online] Available at: https://www.eviews.com/help/helpintro.html#page/content/ardl_Estimating_ARDL_Models_in_EViews.html.

Fettke, K. (2020). *History of the U.S. Housing Market: Great Depression to Donald Trump*. [online] RealWealth. Available at: https://realwealth.com/learn/us-housing-market-history/.

Freese, J. (2015). The Regional Pattern of the U.S. House Price Bubble – An Application of SPC to City Level Data. *Review of Economics*, 66(2). doi:10.1515/roe-2015-0204.

Friedman, N. (2021). WSJ News Exclusive | U.S. Housing Market Needs 5.5 Million More Units, Says New Report. *Wall Street Journal*. [online] 16 Jun. Available at: https://www.wsj.com/articles/u-s-housing-market-needs-5-5-million-more-units-says-new-report-11623835800.

Fritts, J. (2022). *How Does Your State Rank on Property Taxes?* [online] Tax Foundation. Available at: https://taxfoundation.org/ranking-property-taxes-2022/.

Gallin, J. (2006). The Long-Run Relationship between House Prices and Income: Evidence from Local Housing Markets. *Real Estate Economics*, 34(3), pp.417–438. doi:10.1111/j.1540-6229.2006.00172.x.

Glickman, E. (2014). *An introduction to real estate finance*. Waltham, Ma: Academic Press, An Imprint Of Elsevier.

Gros, D. (2007). Bubbles in Real Estate? A Longer-Term Comparative Analysis of Housing Prices in Europe and the Us. *CEPS Working Documents*, [online] No. 239. Available at: https://www.ceps.eu/ceps-publications/bubbles-real-estate-longer-term-comparative-analysis-housing-prices-europe-and-us-0/.

Hiscock, K. (2016). *What Makes Real Estate Markets Different?* [online] www.rochesterrealestateblog.com. Available at: https://www.rochesterrealestateblog.com/what-makes-real-estate-markets-different/.

Hlaváček, M. and Kalabiška, R. (2020). Regional Determinants of Housing Prices in

the Czech Republic. *IES Working Papers 9/2020. IES FSV. Charles University.*.

Hlaváček, M. and Komárek, L. (2009). Housing Price Bubbles and their Determinants in the Czech Republic and its Regions. *Working Papers 2009/12, Czech National Bank.*.

Hlouskova, J. and Wagner, M. (2006). The Performance of Panel Unit Root and Stationarity Tests: Results from a Large Scale Simulation Study. *Econometric Reviews*, 25(1), pp.85–116. doi:10.1080/07474930500545504.

Holly, S., Pesaran, M.H. and Yamagata, T. (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics*, 158(1), pp.160–173. doi:10.1016/j.jeconom.2010.03.040.

Housing Vacancies and Homeownership. (2022). [online] *Census.gov*. Available at: https://www.census.gov/housing/hvs/definitions.pdf.

Im, K.S., Pesaran, M.Hashem. and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, [online] 115(1), pp.53–74. doi:10.1016/s0304-4076(03)00092-7.

Jones, D. (2021). *Age of Houses in the US*. [online] Buyers Ask. Available at: https://buyersask.com/articles-of-interest/age-of-houses-in-the-us-whats-the-average-age/.

Juneja, P. (2021). *The American Real Estate Market*. [online] www.managementstudyguide.com. Available at: https://www.managementstudyguide.com/american-real-estate-market.htm.

Lau, L.-S., Ng, C.-F., Cheah, S.-P. and Choong, C.-K. (2019). Panel Data Analysis (Stationarity, Cointegration, and Causality). *Environmental Kuznets Curve (EKC)*, pp.101–113. doi:10.1016/b978-0-12-816797-7.00009-6.

Levin, A., Lin, C.-F. and James Chu, C.-S. (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics*, 108(1), pp.1–24. doi:10.1016/s0304-4076(01)00098-7.

Liew, V., Khim and Sen (2004). Which Lag Length Selection Criteria Should We Employ? *Economics Bulletin*, [online] 3(33), pp.1–9. Available at: http://accessecon.com/pubs/eb/2004/volume3/EB-04C20021A.pdf.

Logan, L.K. (2022). *Federal Reserve Asset Purchases: The Pandemic Response and Considerations Ahead – Federal Reserve Bank of New York*. [online] www.newyorkfed.org. Available at: https://www.newyorkfed.org/newsevents/speeches/2022/log220302.

Mikhed, V. and Zemčík, P. (2007). Testing for Bubbles in Housing Markets: A Panel Data Approach. *The Journal of Real Estate Finance and Economics*, 38(4), pp.366–386. doi:10.1007/s11146-007-9090-2.

Mourouzi-Sivitanidou, R. (2002). Office Rent Processes: The Case of U.S. Metropolitan Markets. *Real Estate Economics*, [online] 30(2), pp.317–344. doi:10.1111/1540-6229.00042.

Pariona, A. (2017). *US States By Vehicles Per Capita*. [online] WorldAtlas. Available at: https://www.worldatlas.com/articles/us-states-by-vehicles-per-capita.html.

Pedroni, P. (1999). Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors. *Oxford Bulletin of Economics and Statistics*, [online] 61(s1), pp.653–670. doi:10.1111/1468-0084.0610s1653.

Pedroni, P. (2004). Panel cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with Application to the PPP Hypothesis. *Econometric Theory*, [online] 20(03). doi:10.1017/s0266466604203073.

Pesaran, M.H. (1995). An Autoregressive Distributed-Lag Modelling Approach to Cointegration Analysis. *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*, pp.371–413. doi:10.1017/ccol0521633230.011.

Statista. (2021). *Age of U.S. homes 2021*. [online] Available at: https://www.statista.com/statistics/1042458/home-age-usa/ [Accessed 21 Jul. 2022].

Statista. (2022a). *Single-person households United States 2019*. [online] Available at: https://www.statista.com/statistics/242022/number-of-single-person-households-in-the-us/.

Statista. (2022b). *Size of urban and rural population of the U.S. 2019*. [online] Available at: https://www.statista.com/statistics/985183/size-urban-rural-population-us/.

Stlouisfed.org. (2019). *30-Year Fixed Rate Mortgage Average in the United States*.

[online] Available at: https://fred.stlouisfed.org/series/MORTGAGE30US.

The Industry Insights. (2021). *Transaction Costs In The Real Estate Market*. [online] Available at: https://theindustryinsight.com/transaction-costs-in-the-real-estate-market/.

U.S. Census Bureau (2016). *Understanding Population Density*. [online] U.S. Census Bureau. Available at: https://www.census.gov/newsroom/blogs/randomsamplings/2015/03/understanding-population-density.html.

U.S. Census Bureau. (2019). *American Housing Survey (AHS)*. [online] Available at: https://www.census.gov/programs-surveys/ahs/data/interactive/ahstablecreator.html

U.S. Census Bureau (2022a). *Average Number of Months from Start to Completion*. [online] U.S. Census Bureau. Available at: https://www.census.gov/construction/nrc/pdf/avg_starttocomp.pdf.

U.S. Census Bureau QuickFacts. (2021). *U.S. Census Bureau QuickFacts*. [online] Available at: https://www.census.gov/quickfacts/fact/table/NJ.

U.S. Census Bureau, M.C.D. (2022b). *U.S. Census Bureau Building Permits Survey*. [online] www.census.gov. Available at: https://www.census.gov/construction/bps/.

U.S. Census Bureau. (2020). *Census profile: New York-Newark-Jersey City, NY-NJ-PA Metro Area*. [online] Available at:https://censusreporter.org/profiles/31,000US35620-new-york-newark-jersey-city-ny-nj-pa-metro-area/.

Weinstock, L.R. (2021). Introduction to U.S. Economy: Housing Market. [online] Available at: https://crsreports.congress.gov/product/pdf/IF/IF11327.

York, E. and Walczak, J. (2022). *State and Local Tax Burdens, Calendar Year 2022*. [online] Tax Foundation. Available at: https://taxfoundation.org/tax-burden-by-state-2022/.

Zhao, Y. (2020). US Housing Market During COVID-19: Aggregate and Distributional Evidence. *SSRN Electronic Journal*, 50: 113-154. doi:10.2139/ssrn.3744679.

Zillow Research. (2022). *Housing Data - Zillow Research*. [online] Available at: https://www.zillow.com/research/data/.

Appendix

Table A.1. Summary statistics

Variable	Units	Mean	Median	Standard deviation	Variability over time[a)]	Variability across States[a)]	Minimum across States	Maximum across States
Home value	dollars	204849.55	178562.83	96892.91	17.38%	42.23%	West Virginia	Hawaii
Housing units	per 1,000 inhabitants	436.60	436.91	32.73	1.73%	7.43%	Utah	Maine
Building permits	per 1,000 inhabitants	0.36	0.30	0.24	44.94%	51.85%	D.C.	Nevada
Construction value per unit	thousands of dollars	165.96	163.00	43.45	20.16%	18.63%	D.C.	Hawaii
Homeowner vacancy rates	percentage	1.94	1.90	0.80	35.61%	35.51%	Utah	Nevada
Homeownership rates	percentage	68.22	69.35	6.32	3.51%	9.02%	D.C.	West Virginia
Rental vacancy rates	percentage	8.68	8.30	3.18	26.28%	34.68%	Maine	Mississippi
Personal income per capita	dollars	41728.05	40378.50	10628.72	19.50%	16.56%	Mississippi	D.C.
PCE per capita	dollars	34142.89	33600.12	7712.64	16.93%	15.21%	Mississippi	D.C.
Unemployment rate	percentage	5.60	5.17	2.17	32.27%	22.93%	Hawaii	Nevada
Tax burden rate	percentage	9.79	9.80	1.38	5.39%	13.72%	Alaska	New York
Debt-to-income ratio	ratio	1.61	1.56	0.41	11.73%	22.33%	D.C.	California
Population density	citizens per km2	151.14	39.39	548.51	4.82%	364.56%	Alaska	D.C.
Share of population 20-39	per 1,000 inhabitants	271.31	268.12	21.38	2.45%	7.74%	Maine	D.C.
Share of population 15-64	per 1,000 inhabitants	664.2	664.56	17.42	1.43%	2.28%	South Dakota	D.C.
Population growth	qt percentage change	0.18	0.16	0.19	206.12%	104.82%	Louisiana	Nevada
Marriage rates	per 1,000 inhabitants	7.93	6.90	5.90	12.00%	66.96%	California	Nevada
30-year fixed mortgage rate	percentage	5.00	4.88	1.32	26.51%	-	-	-

a) Variability computed as standard deviation in % of mean

Table A.2. Unit root tests (IPS, LLC and Breitung)

First-generation	Im-Pesaran-Shin		Levin-Lin-Chu		Breitung	
Unit root tests	statistic	p-value	statistic	p-value	statistic	p-value
Home value	18.848	1	7.469	1	31.595	1
Housing units	-9.549	0***	-2.166	0.015**	23.706	1
Building permits	-14.326	0***	-6.270	0***	-13.508	0***
Construction value per unit	-7.678	0***	-5.598	0***	-1.761	0.039**
Homeowner vacancy rates	-22.045	0***	-13.314	0***	-20.927	0***
Homeownership rates	-15.829	0***	-10.783	0***	-15.797	0***
Rental vacancy rates	-18.773	0***	-10.569	0***	-18.101	0***
Personal income per capita	15.111	1	8.504	1	22.863	1
PCE per capita	-8.933	0***	-12.485	0***	34.572	1
Unemployment rate	-10.601	0***	-4.739	0***	-11.094	0***
Tax burden	2.752	0.997	3.115	0.999	1.549	0.939
Debt-to-income ratio	-5.181	0***	-4.165	0***	-2.264	0.012**
Population density	-3.940	0***	-9.699	0***	31.630	1
Share of population 20-39	-29.868	0***	-9.539	0***	16.747	1
Share of population 15-64	5.730	1	22.775	1	22.976	1
Population growth	-13.309	0***	-5.856	0***	-11.195	0***
Marriage rates	5.906	1	0.800	0.788	8.608	1
30-year mortgage rate	-4.680	0***	-6.606	0***	-4.419	0***
log(home value)	5.161	1	1.673	0.953	28.795	1
log(construction)	-8.137	0***	-7.695	0***	-0.686	0.246
log(income)	8.437	1	-0.749	0.227	24.258	1
log(pce)	-22.780	0***	-14.295	0***	34.151	1

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level

Table A.3. Unit root tests (ADF and PP)

Augmented Dickey-Fuller/ Phillips-Perron	ADF Fisher-type		ADF Choi		PP Fisher-type		PP Choi	
	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value
Home value	63.246	1	6.196	1	42.832	1	7.052	1
Housing units	-9.549	0***	-2.166	0.015**	585.296	0***	-10.079	0***
Building permits	-14.326	0***	-6.270	0***	506.077	0***	-14.709	0***
Construction value per unit	-7.678	0***	-5.598	0***	221.607	0***	-5.313	0***
Homeowner vacancy rates	-22.045	0***	-13.314	0***	58.449	0***	-24.766	0***
Homeownership rates	-15.829	0***	-10.783	0***	31.111	0***	-16.693	0***
Rental vacancy rates	-18.773	0***	-10.569	0***	44.678	0***	-20.496	0***
Personal income per capita	15.111	1	8.504	1	4.977	1	14.237	1
PCE per capita	-8.933	0***	-12.485	0***	172.902	0***	-4.683	0***
Unemployment rate	-10.601	0***	-4.739	0***	309.269	0***	-10.974	0***
Tax burden	2.752	0.997	3.115	0.999	77.831	0.951	2.514	0.994
Debt-to-income ratio	-5.181	0***	-4.165	0***	186.350	0***	-5.190	0***
Population density	-3.940	0***	-9.699	0***	440.936	0***	-5.814	0***
Share of population 20-39	-29.868	0***	-9.539	0***	109.524	0***	-30.346	0***
Share of population 15-64	5.730	1	22.775	1	2.429	1	8.589	1
Population growth	-13.309	0***	-5.856	0***	407.719	0***	-12.844	0***
Marriage rates	5.906	1	0.800	0.788	42.564	1	5.762	1
30-year fixed mortgage rate	-4.680	0***	-6.606	0***	158.006	0***	-5.635	0***
log(home value)	5.161	1	1.673	0.953	156.290	0***	2.235	0.987
log(construction)	-8.137	0***	-7.695	0***	222.281	0***	-6.334	0***
log(income)	8.437	1	-0.749	0.227	23.050	1	8.314	1
log(pce)	95.387	0.612	3.535	1	3.980	0***	2.235	0.987

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level

Table A.4. Second generation unit root tests (CIPS and CADF)

Second-generation tests	CIPS	CADF
	statistic	statistic
Home value	-1.079	-1.686
Housing units	-0.709	-1.071
Building permits	-5.150***	-3.849***
Construction value per unit	-5.749***	-4.226***
Homeowner vacancy rates	-5.566***	-4.636***
Homeownership rates	-5.429***	-4.777***
Rental vacancy rates	-5.183***	-4.457***
Personal income per capita	-1.898	-1.686
PCE per capita	-0.159	-1.765
Unemployment rate	-2.712***	-2.464**
Tax burden	-1.818	-1.872
Debt-to-income ratio	-2.670***	-2.434**
Population density	-0.980	-0.706
Share of population 20-39	-1.343	-1.029
Share of population 15-64	-1.114	-1.403
Population growth	-3.964***	-2.990***
Marriage rates	-2.089*	-2.211*
log(home value)	-1.704	-1.92
log(valuation)	-5.739***	-4.283***
log(income)	-1.894	-1.761
log(pce)	-0.581	-1.815

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level
Critical values for CIPS: -2.23 for 1 % significance level, -2.12 for 5 % significance level, -2.05 for 10 % significance level

Table A.5. Unit root tests on first differences (1)

First differences	Im-Pesaran-Shin		Levin-Lin-Chu		Breitung	
	statistic	p-value	statistic	p-value	statistic	p-value
Home value	1.369	0.915	4.379	1.000	-7.470	0***
Personal income per capita	-46.809	0***	-29.409	0***	-46.672	0***
Tax burden	-42.366	0***	-36.542	0***	-45.000	0***
Share of population 15-64	-16.378	0***	-11.960	0***	-12.230	0***
Marriage rates	-43.195	0***	-38.858	0***	-45.000	0***
log(home value)	-1.846	0.033**	0.045	0.051*	-9.303	0***
log(income)	-45.231	0***	-26.945	0***	-40.223	0***
Additional variables from Breitung, CIPS and CADF test						
Housing units	-12.274	0***	-8.162	0***	-6.167	0***
PCE per capita	-2.271	0.012**	2.529	0.994	-11.748	0***
Population density	-35.752	0***	-14.172	0***	-20.624	0***
Share of population 20-39	-12.021	0***	-8.064	0***	-4.464	0***
log(construction)	-52.814	0***	-65.367	0***	-43.015	0***
log(pce)	-1.524	0.064*	1.712	0.957	-8.312	0***

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level

Table A.6. Unit root tests on first differences (2)

First differences	ADF Fisher-type		ADF Choi		PP Fisher-type		PP Choi	
	statistic	p-val.	statistic	p-val.	statistic	p-val.	statistic	p-val.
Home value	1.060	0.145	-0.532	0.297	1.104	0.135	0.131	0.552
Personal income per capita	131.570	0***	-40.509	0***	239.958	0***	-56.460	0***
Tax burden	118.228	0***	-38.805	0***	227.434	0***	-54.931	0***
Share of population 15-64	11.585	0***	-9.122	0***	29.444	0***	-16.134	0***
Marriage rates	131.730	0***	-41.033	0***	236.246	0***	-56.030	0***
log(home value)	3.575	0***	-3.709	0***	3.567	0***	-2.884	0.002***
log(income)	111.335	0***	-37.058	0***	232.907	0***	-55.518	0***
Additional variables from Breitung, CIPS and CADF test								
Housing units	19.275	0***	-10.470	0***	20.558	0***	-12.357	0***
PCE per capita	0.464	0.321	-2.249	0.012**	1.109	0.134	-2.616	0.004***
Population density	87.712	0***	-27.594	0***	162.213	0***	-41.461	0***
Share of population 20-39	8.985	0***	-6.391	0***	21.481	0***	-11.234	0***
log(construction)	240.241	0***	-56.504	0***	247.796	0***	-57.459	0***
log(pce)	3.575	0***	-3.709	0***	3.567	0***	-2.884	0.002***

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level

Table A.7. Correlation matrix

	Housing units	Building permits	Construction value per unit	Homeowner vacancy rates	Homeownership rates	Rental vacancy rates	Personal income per capita	PCE per capita	Unemployment rate	Tax burden	Debt-to-income ratio	Population density	Share of population 20-39	Share of population 15-64	Population growth	Marriage rates	30-year mortgage rate
Housing units	1.00																
Building permits	-0.16	1.00															
Construction value	-0.05	-0.28	1.00														
Homeowner vacancy rates	0.08	0.03	-0.23	1.00													
Homeownership rates	0.25	0.10	-0.15	0.06	1.00												
Rental vacancy rates	0.00	0.12	-0.30	0.52	0.26	1.00											
Personal income	0.01	-0.26	0.56	-0.32	-0.47	-0.44	1.00										
PCE per capita	0.16	-0.26	0.54	-0.27	-0.49	-0.44	0.95	1.00									
Unemployment rate	0.01	-0.30	0.01	0.25	-0.13	0.21	0.01	0.01	1.00								
Tax burden	-0.04	-0.26	0.13	-0.16	-0.25	-0.31	0.36	0.34	0.05	1.00							
Debt-to-income ratio	-0.20	0.13	0.30	0.20	0.11	0.02	-0.09	-0.05	0.14	0.00	1.00						
Population density	0.13	-0.02	-0.17	0.02	-0.59	-0.04	0.37	0.47	0.12	0.17	-0.36	1.00					
Share of population 20-39	-0.38	0.27	-0.14	0.01	-0.60	0.03	0.17	0.18	0.05	-0.08	-0.22	0.69	1.00				
Share of population 15-64	0.02	0.00	-0.23	0.14	-0.36	-0.02	0.07	0.16	0.24	-0.04	0.02	0.54	0.49	1.00			
Population growth	-0.23	0.62	-0.14	0.11	-0.05	0.08	-0.21	-0.18	-0.10	-0.31	0.26	0.03	0.32	0.14	1.00		
Marriage rates	-0.11	0.37	-0.09	0.18	-0.15	0.10	-0.16	-0.14	-0.02	-0.20	0.20	-0.05	0.12	0.07	0.37	1.00	
30-year mortgage rate	0.00	-0.03	-0.03	0.03	0.01	0.03	-0.05	-0.04	0.03	0.00	0.06	0.01	0.00	0.06	0.03	0.00	1.00

Table A.8. Granger causality test

Granger causality	H0: … does not Granger cause home value		H0: **home value** *does not Granger cause …*	
	F-stats	p-value	F-stats	p-value
Housing units	1.619	0.198	24.766	0***
Δ Housing units	3.933	0.020**	23.766	0***
Building permits	23.135	0***	18.704	0***
Construction value per unit	2.087	0.124	4.463	0.012**
Homeowner vacancy rates	51.503	0***	37.457	0***
Homeownership rates	2.168	0.0114**	2.340	0.097*
Rental vacancy rates	10.742	0***	10.274	0***
Δ Log (personal income)	33.553	0***	50.621	0***
Unemployment rate	36.967	0***	107.083	0***
Δ Tax burden	0.133	0.876	4.707	0.009***
Debt-to-income ratio	15.597	0***	130.743	0***
Population density	0.552	0.576	2.741	0.065
Δ Population density	1.573	0.208	3.114	0.045**
Share of population 20-39	1.054	0.347	5.872	0.003***
Δ Share of population 20-39	0.228	0.796	4.753	0.009***
Δ Share of population 15-64	6.638	0.001***	18.829	0***
Population growth	3.259	0.039**	11.854	0***
Δ Marriage rates	3.838	0.022**	2.084	0.125
S&P 500	57.717	0***	12.967	0***
30-year fixed mortgage rate	3.4996	0.030**	0.794	0.452

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level

Table A.9. An overview of U.S. regions and divisions

Northeast	Midwest	South	West
New England (5)	East North Central (1)	West South Central (9)	Pacific (6)
Connecticut Maine Massachusetts New Hampshire Rhode Island Vermont	Illinois Indiana Michigan Ohio Wisconsin	Arkansas Louisiana Oklahoma Texas West	Alaska California Hawaii Oregon Washington
Mid-Atlantic (4)	West North Central (8)	East South Central (2)	Mountain (3)
New Jersey New York Pennsylvania	Iowa Kansas Minnesota Missouri Nebraska North Dakota South Dakota	Alabama Kentucky Mississippi Tennessee	Arizona Colorado Idaho Montana Nevada New Mexico Utah Wyoming
		South Atlantic (7)	
		Delaware Florida Georgia Maryland North Carolina South Carolina Virginia Washington West Virginia	

Table A.10. Unit root tests by division– divisions (1) to (3)

Im-Pesaran-Shin and CIPS	(1) IPS		(1) CIPS	(2) IPS		(2) CIPS	(3) IPS		(3) CIPS
Unit root test	statistic	p-value	statistic	statistic	p-value	statistic	statistic	p-value	statistic
Home values	2.672	1	-0.435	2.945	1	-0.163	2.014	1	-1.887
Housing units	-7.238	0***	-1.461	-6.340	0***	-1.015	-3.548	0***	-0.297
Building permits	-2.930	0***	-5.568***	-2.661	0.005***	-2.994***	-2.999	0***	-2.262*
Construction value per unit	-1.890	0.182	-2.996***	-1.542	0.468	-5.394***	-2.391	0.002***	-4.108***
Homeowner vacancy rates	-3.759	0***	-5.147***	-3.759	0***	-6.004***	-4.436	0***	-5.094***
Homeownership rates	-2.964	0***	-5.369***	-3.558	0***	-4.760***	-3.888	0***	-4.650***
Rental vacancy rates	-3.464	0***	-5.525***	-4.890	0***	-6.080***	-3.815	0***	-3.206***
Personal income per capita	0.727	1	-1.978	-0.046	1	-0.862	0.350	1	-1.270
PCE per capita	-2.168	0.047**	-2.349***	-2.477	0.018**	-2.126	-1.985	0.085*	-1.596
Unemployment rate	-3.064	0***	-2.586***	-2.591	0.007***	-2.948***	-2.319	0.005***	-2.281*
Tax burden	-1.874	0.189	-1.333	-1.536	0.480	-1.469	-1.120	0.911	-2.066
Debt-to-income ratio	-1.492	0.522	-2.585***	-3.454	0***	-4.446***	-1.987	0.060*	-2.583***
Population density	-2.040	0.093*	0.206	-2.303	0.040**	0.898	-0.957	0.972	-2.156
Share of population 20-39	-10.716	0***	-0.327	-10.435	0***	0.052	-3.183	0***	-1.167
Share of population 15-64	5.645	1	-2.247*	6.987	1	-2.317*	3.406	1	-2.687***
Population growth	-3.926	0***	-3.226***	-3.438	0***	-5.361***	-2.986	0***	-2.983***
Marriage rates	-0.831	0.967	-2.767***	-1.431	0.597	-1.057	-1.155	0.887	-2.339**
log(home values)	1.490	1	-0.958	1.169	1	-0.381	0.077	1	-2.421**
log(income)	0.128	1	-2.423*	-0.737	0.968	-1.137	-0.442	1	-1.500
log(pce)	-4.191	0***	-2.349**	-4.980	0***	-2.237*	-3.775	0***	-1.664

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level

Critical values for CIPS: -2.53 for 1 % significance level, -2.32 for 5 % significance level, -2.21 for 10 % significance level

Table A.11. Unit root tests by division – divisions (4) to (6)

Im-Pesaran-Shin	(4) IPS		(4) CIPS	(5) IPS		(5) CIPS	(6) IPS		(6) CIPS
Unit root test	statistic	p-value	statistic	statistic	p-value	statistic	statistic	p-value	statistic
Home values	-2.479	0.018**	-0.612	-1.632	0.388	1.098	-0.283	1	-1.646
Housing units	0.404	1	-0.719	-1.986	0.210	-1.266	1.000	0.983	-0.665
Building permits	-5.218	0***	-5.667***	-3.995	0***	-4.609***	-3.275	0***	-4.475***
Construction value per unit	-3.416	0***	-4.044***	-3.574	0***	-5.563***	-2.688	0.002***	-5.629***
Homeowner vacancy rates	-5.624	0***	-6.190***	-5.548	0***	-6.035***	-4.696	0***	-5.188***
Homeownership rates	-3.731	0***	-4.099***	-3.539	0***	-4.232***	-3.823	0***	-4.751***
Rental vacancy rates	-4.611	0***	-5.149***	-4.587	0***	-5.501***	-4.072	0***	-3.139***
Personal income per capita	0.268	1	-2.474**	0.278	1	-1.937	0.791	1	-0.877
PCE per capita	-3.327	0***	-2.131	-3.477	0***	-1.573	-2.188	0.051*	-1.346
Unemployment rate	-3.414	0***	-1.776	-2.964	0***	-1.536	-3.273	0***	-1.086
Tax burden	-1.315	0.690	-2.109	-1.291	0.718	-1.928	-0.434	0.998	-1.453
Debt-to-income ratio	-2.311	0.033**	-2.186	-2.211	0.036**	-2.540***	-1.649	0.372	-0.879
Population density	-5.313	0***	-1.508	-3.498	0***	-1.783	-2.511	0.073*	-1.098
Share of population 20-39	-9.581	0***	-0.788	-6.833	0***	-2.445**	-2.700	0.030**	-0.660
Share of population 15-64	3.516	1	0.151	1.587	1	0.534	3.659	1	-2.037
Population growth	-2.719	0***	-3.917***	-3.118	0***	-2.730***	-2.714	0.002***	-3.601***
Marriage rates	-0.608	0.985	-2.139	-0.956	0.934	-2.690***	2.060	1	-2.243*
log(home values)	-4.733	0***	-1.387	-3.622	0***	0.681	-2.116	0.134	-2.221*
log(income)	-0.571	0.988	-2.764***	-0.497	0.997	-0.033	-0.133	1	-1.228
log(pce)	-5.779	0***	-2.267*	-6.035	0***	-1.516	-3.910	0***	-1.458

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level

Critical values for CIPS: -2.53 for 1 % significance level, -2.32 for 5 % significance level, -2.21 for 10 % significance level

Table A.12. Unit root tests by division – divisions (7) to (9)

Im-Pesaran-Shin	(7) (without D.C.)		(7) CIPS	(8) IPS		(8) CIPS	(9) IPS		(9) CIPS
Unit root test	statistic	p-value	statistic	statistic	p-value	statistic	statistic	p-value	statistic
Home values	4.900	0.998	-1.289	2.429	1	-2.118	1.751	1	-1.652
Housing units	-3.693	0***	-1.843	-3.827	0***	-1.657	-2.659	0.017**	-1.179
Building permits	-2.019	0.053*	-3.915***	-4.590	0***	-5.501***	-3.594	0***	-5.524***
Construction value per unit	-2.137	0.022**	-3.594***	-2.505	0.003***	-5.027***	-1.762	0.283	-4.602***
Homeowner vacancy rates	-4.287	0***	-5.795***	-4.117	0***	-4.941***	-4.872	0***	-4.918***
Homeownership rates	-3.222	0***	-4.603***	-4.127	0***	-4.637***	-3.889	0***	-3.928***
Rental vacancy rates	-3.314	0***	-4.772***	-4.983	0***	-5.804***	-3.950	0***	-4.274***
Personal income per capita	5.546	0.998	-1.955	0.088	1	-2.024	-0.507	0.992	-1.620
PCE per capita	-2.790	0***	-1.347	-3.130	0***	-2.125	-2.595	0.007***	-1.488
Unemployment rate	-2.565	0***	-1.573	-3.161	0***	-1.521	-3.192	0***	-2.892***
Tax burden	-1.013	0.955	-1.104	-0.753	0.988	-1.500	-1.759	0.279	-1.844
Debt-to-income ratio	-2.064	0.039**	-2.238*	-2.264	0.017**	-2.906***	-2.684	0.004***	-1.995
Population density	-4.365	0***	0.132	-1.046	0.920	-2.331**	-0.966	0.904	1.233
Share of population 20-39	-13.001	0***	-2.649***	-6.620	0***	-2.629***	-5.797	0***	-3.734***
Share of population 15-64	5.811	1	-1.628	5.203	1	-1.739	4.323	1	-1.025
Population growth	-2.810	0.003***	-2.645***	-3.815	0***	-4.121***	-4.232	0***	-1.769
Marriage rates	2.060	0.980	-2.666***	-0.492	0.999	-2.240*	-0.941	0.913	-2.506**
log(home values)	0.740	0.770	-1.870	0.336	1	-2.859***	-0.407	0.997	-1.607
log(income)	-0.614	0.999	-2.163	-0.671	0.993	-2.235*	-1.265	0.719	-1.634
log(pce)	-4.812	0***	-1.509	-5.456	0***	-2.285*	-4.689	0***	-1.422

*** - significant at 1% level, ** - significant at 5% level, * - significant at 10% level

Critical values for CIPS: -2.53 for 1 % significance level, -2.32 for 5 % significance level, -2.21 for 10 % significance level

Table A.13. Pedroni test for cointegration by division

Pedroni test for cointegration	(1)		(2)		(3)	
Modified Phillips–Perron t	3.9528	0.0000***	1.6152	0.0531**	1.9772	0.0240**
Phillips–Perron t	5.8981	0.0000***	1.4132	0.0788*	1.8157	0.0347**
Augmented Dickey–Fuller t	8.1579	0.0000***	2.5573	0.0053***	2.5335	0.0056***

Pedroni test for cointegration	(4)		(5)		(6)	
Modified Phillips–Perron t	3.0406	0.0012***	2.9398	0.0016***	2.0157	0.0219**
Phillips–Perron t	3.3523	0.0004***	2.0806	0.0187**	1.9612	0.0249**
Augmented Dickey–Fuller t	3.8323	0.0001***	2.3166	0.0103**	1.6082	0.0539*

Pedroni test for cointegration	(7) without D.C.		(8)		(9)	
Modified Phillips–Perron t	2.6364	0.0042***	2.2603	0.0119**	1.8005	0.0359**
Phillips–Perron t	2.3385	0.0097***	1.7526	0.0398**	1.4418	0.0747*
Augmented Dickey–Fuller t	2.0318	0.0211**	1.9736	0.0242**	2.1651	0.0152**

Table A.14. Summary statistics for each division - divisions (1) and (2)

	Division 1: East North Central				
Variable	Obs	Mean	Std. Dev.	Min	Max
Home values	420	147890.90	32121.47	95963.30	239135.00
Housing units	420	437.47	18.71	390.70	469.20
Building permits	420	0.26	0.15	0.00	0.70
Construction value per unit	420	172.33	33.73	109.20	256.00
Homeowner vacancy rates	420	2.08	0.82	0.30	4.40
Homeownership rates	420	71.02	3.31	63.40	78.20
Rental vacancy rates	420	9.76	3.12	3.00	19.40
Personal income per capita	420	40000.65	7733.21	27770.00	63808.00
PCE per capita	420	32699.20	5713.69	21967.90	45257.10
Unemployment rate	420	6.27	2.35	2.90	18.80
Tax burden	420	10.13	0.78	8.30	12.50
Debt-to-income ratio	420	1.42	0.20	1.00	1.90
Population density	420	75.08	22.96	38.30	110.60
Share of population 20-39	420	264.53	9.98	247.30	294.70
Share of population 15-64	420	661.86	8.30	637.00	675.20
Population growth	420	0.05	0.08	-0.20	0.20
Marriage rates	420	6.12	0.77	3.90	8.00
	Division 2: East South Central				
Variable	Obs	Mean	Std. Dev.	Min	Max
Home values	336	124811.20	21528.47	83762.00	212882.30
Housing units	336	441.07	11.73	401.70	467.20
Building permits	336	0.31	0.15	0.10	0.70
Construction value per unit	336	141.47	30.61	71.90	216.00
Homeowner vacancy rates	336	2.10	0.68	0.40	4.00
Homeownership rates	336	71.86	2.94	64.30	80.40
Rental vacancy rates	336	11.17	2.95	4.30	19.70
Personal income per capita	336	34190.08	6589.54	21294.90	52674.00
PCE per capita	336	28520.82	4913.89	17710.60	38292.10
Unemployment rate	336	6.35	2.03	2.90	12.70
Tax burden	336	8.83	0.95	6.90	10.20
Debt-to-income ratio	336	1.47	0.13	0.76	1.75
Population density	336	40.37	12.77	23.40	64.70
Share of population 20-39	336	268.24	7.31	259.10	290.80
Share of population 15-64	336	661.45	9.98	635.80	675.40
Population growth	336	0.13	0.10	-0.20	0.40
Marriage rates	336	7.98	1.92	4.80	15.50

Table A.15. Summary statistics for each division - divisions (3) and (4)

	Division 3: Mountain				
Variable	Obs	Mean	Std. Dev.	Min	Max
Home values	672	215541.60	59176.50	126225.00	436925.00
Housing units	672	429.11	36.70	338.90	486.70
Building permits	672	0.51	0.33	0.10	2.00
Construction value per unit	672	178.63	47.14	77.80	401.50
Homeowner vacancy rates	672	2.04	0.93	0.10	6.00
Homeownership rates	672	68.66	4.65	52.90	77.60
Rental vacancy rates	672	8.16	3.03	2.40	18.90
Personal income per capita	672	38687.59	9283.94	22436.30	66212.00
PCE per capita	672	31947.60	6049.21	20053.00	47666.40
Unemployment rate	672	5.42	2.25	2.40	23.50
Tax burden	672	9.15	0.98	6.30	11.80
Debt-to-income ratio	672	1.90	0.32	1.17	2.85
Population density	672	10.14	6.80	2.00	25.40
Share of population 20-39	672	276.29	17.90	243.90	315.00
Share of population 15-64	672	658.16	17.05	624.80	693.60
Population growth	672	0.36	0.22	-0.20	1.20
Marriage rates	672	12.18	13.14	3.40	72.20
	Division 4: Mid-Atlantic				
Variable	Obs	Mean	Std. Dev.	Min	Max
Home values	336	235244.60	67895.82	106580.70	403550.70
Housing units	336	442.86	44.53	390.70	547.80
Building permits	336	0.22	0.10	0.10	0.70
Construction value per unit	336	153.46	31.87	90.50	278.40
Homeowner vacancy rates	336	1.65	0.54	0.30	3.20
Homeownership rates	336	65.98	7.77	50.20	76.80
Rental vacancy rates	336	6.27	2.16	1.80	13.70
Personal income per capita	336	47156.78	10123.26	28254.60	74159.00
PCE per capita	336	38430.88	7140.35	24620.90	52063.50
Unemployment rate	336	5.53	2.05	2.10	15.50
Tax burden	336	11.59	1.24	9.80	14.90
Debt-to-income ratio	336	1.45	0.28	0.81	2.06
Population density	336	188.24	163.81	25.50	467.00
Share of population 20-39	336	260.05	13.56	238.40	293.60
Share of population 15-64	336	667.66	12.41	637.60	692.30
Population growth	336	0.03	0.07	-0.20	0.20
Marriage rates	336	6.62	1.42	4.10	10.00

Table A.16. Summary statistics for each division - divisions (5) and (6)

	Division 5: Nex England				
Variable	Obs	Mean	Std. Dev.	Min	Max
Home values	420	259625.80	65660.97	116891.00	467954.00
Housing units	420	453.92	45.80	405.70	558.80
Building permits	420	0.21	0.13	0.00	0.80
Construction value per unit	420	185.01	37.66	93.70	387.20
Homeowner vacancy rates	420	1.53	0.60	0.20	3.20
Homeownership rates	420	68.14	5.54	54.40	79.30
Rental vacancy rates	420	5.86	1.94	1.50	11.60
Personal income per capita	420	48820.19	11924.38	26894.80	80278.00
PCE per capita	420	39410.94	7271.89	24171.90	54787.10
Unemployment rate	420	5.41	2.25	2.00	15.80
Tax burden	420	10.68	1.30	7.60	13.70
Debt-to-income ratio	420	1.66	0.26	0.99	2.19
Population density	420	215.63	151.30	16.00	401.90
Share of population 20-39	420	255.89	16.27	227.00	295.00
Share of population 15-64	420	671.72	11.45	630.60	689.20
Population growth	420	0.07	0.10	-0.20	0.40
Marriage rates	420	6.54	1.09	4.00	9.40
	Division 6: Pacific				
Variable	Obs	Mean	Std. Dev.	Min	Max
Home values	420	332409.10	132943.70	130847.30	678968.00
Housing units	420	404.15	28.14	357.40	440.20
Building permits	420	0.33	0.17	0.00	0.80
Construction value per unit	420	206.94	51.42	107.20	499.90
Homeowner vacancy rates	420	1.63	0.68	0.20	4.00
Homeownership rates	420	62.08	4.45	52.40	70.10
Rental vacancy rates	420	6.42	2.10	2.60	15.80
Personal income per capita	420	44372.91	9933.23	27800.40	72170.00
PCE per capita	420	36480.25	6835.59	23655.40	51417.00
Unemployment rate	420	6.30	2.37	1.90	20.30
Tax burden	420	9.69	2.13	4.90	13.70
Debt-to-income ratio	420	2.04	0.40	1.36	3.15
Population density	420	45.69	36.00	0.40	97.80
Share of population 20-39	420	283.26	10.11	267.50	307.90
Share of population 15-64	420	674.14	16.32	624.10	710.00
Population growth	420	0.22	0.15	-0.20	0.50
Marriage rates	420	8.97	4.84	3.20	22.60

Table A.17. Summary statistics for each division - divisions (7) and (8)

	Division 7: South Atlantic (without D.C.)				
Variable	Obs	Mean	Std. Dev.	Min	Max
Home values	672	190475.70	65211.14	67311.33	369555.00
Housing units	672	440.52	25.19	398.11	503.26
Building permits	672	0.46	0.27	0.05	1.42
Construction value per unit	672	155.06	35.71	90.57	244.52
Homeowner vacancy rates	672	2.25	0.85	0.30	5.40
Homeownership rates	672	71.12	4.30	61.30	82.40
Rental vacancy rates	672	10.12	2.91	3.00	19.20
Personal income per capita	672	40337.03	8974.51	21820.04	67493.00
PCE per capita	672	33492.24	6326.83	19476.77	47128.93
Unemployment rate	672	5779.00	2.14	2.13	13.03
Tax burden	672	9.75	0.83	8.20	12.60
Debt-to-income ratio	672	1.84	0.37	1.09	2.99
Population density	672	106.09	64.50	28.62	241.02
Share of population 20-39	672	264.89	14.53	239.47	316.44
Share of population 15-64	672	664.95	15.91	623.80	689.82
Population growth	672	0.25	0.16	-0.21	0.75
Marriage rates	672	6.94	1.05	4.31	10.60
	Division 8: West North Central				
Variable	Obs	Mean	Std. Dev.	Min	Max
Home values	504	148442.50	35804.47	94296.30	277944.30
Housing units	504	438.87	8.86	417.40	460.10
Building permits	504	0.35	0.15	0.10	0.80
Construction value per unit	504	157.56	33.73	78.80	254.90
Homeowner vacancy rates	504	1.86	0.74	0.40	3.90
Homeownership rates	504	70.56	3.04	60.00	78.80
Rental vacancy rates	504	8.93	2.56	2.10	17.40
Personal income per capita	504	41201.27	8501.20	26127.50	64787.00
PCE per capita	504	32992.35	5871.06	20591.50	46627.50
Unemployment rate	504	4.41	1.47	2.40	9.70
Tax burden	504	9.69	1.13	7.10	12.70
Debt-to-income ratio	504	1.36	0.20	0.76	1.90
Population density	504	17.81	9.99	3.80	34.60
Share of population 20-39	504	264.65	5.74	253.60	286.50
Share of population 15-64	504	651.74	11.16	617.70	674.80
Population growth	504	0.14	0.08	0.00	0.30
Marriage rates	504	6.62	0.85	4.40	9.40

Table A.18. Summary statistics for each division – division (9) and D.C.

	Division 9: West South Central				
Variable	Obs	Mean	Std. Dev.	Min	Max
Home values	336	129212.10	30140.78	79745.00	228852.70
Housing units	336	428.07	23.88	383.60	461.60
Building permits	336	0.35	0.15	0.10	0.90
Construction value per unit	336	149.05	32.17	86.50	217.70
Homeowner vacancy rates	336	2.17	0.70	0.60	4.70
Homeownership rates	336	67.57	3.12	60.90	74.60
Rental vacancy rates	336	10.99	2.23	5.10	16.30
Personal income per capita	336	37411.37	8194.21	22335.50	57213.00
PCE per capita	336	29285.65	5401.27	18958.70	40681.10
Unemployment rate	336	5.44	1.48	2.90	12.30
Tax burden	336	8.94	0.90	7.30	10.90
Debt-to-income ratio	336	1.31	0.11	1.00	1.54
Population density	336	30.10	9.23	19.40	43.60
Share of population 20-39	336	276.11	10.66	260.80	306.40
Share of population 15-64	336	658.08	10.76	630.60	675.10
Population growth	336	0.19	0.31	-3.50	1.80
Marriage rates	336	8.05	2.28	3.40	15.40
	District Of Columbia				
Variable	Obs	Mean	Std. Dev.	Min	Max
Home values	84	446983.80	127017.80	194004.30	656527.00
Housing units	84	476.83	19.97	449.61	506.53
Building permits	84	0.39	0.30	0.02	1.33
Construction value per unit	84	111.02	35.62	44.63	255.72
Homeowner vacancy rates	84	2.23	0.89	0.80	4.30
Homeownership rates	84	43.40	2.37	38.20	48.10
Rental vacancy rates	84	8.50	1.97	5.10	12.50
Personal income per capita	84	64092.28	13840.87	42727.96	91917.00
PCE per capita	84	56116.48	8396.94	39196.12	69780.81
Unemployment rate	84	7.34	1.52	5.00	10.60
Tax burden	84	10.71	0.71	9.60	11.80
Debt-to-income ratio	84	0.55	0.05	0.44	0.66
Population density	84	3950.52	344.85	3583.86	4518.39
Share of population 20-39	84	376.22	21.01	345.66	399.35
Share of population 15-64	84	727.88	12.45	707.10	748.61
Population growth	84	0.27	0.25	-0.38	0.64
Marriage rates	84	6.74	2.21	4.00	11.80