BI

WISEflow
Europe/Oslo(CEST)
30 Jun 2022

Handelshøyskolen BI

GRA 19703 Master Thesis

Thesis Master of Science 100% - W

Predefinert informasjon

Startdato:	16-01-2022 09:00	**Termin:**	202210
Sluttdato:	01-07-2022 12:00	**Vurderingsform:**	Norsk 6-trinns skala (A-F)
Eksamensform:	T		
Flowkode:	202210||10936||IN00||W||T		
Intern sensor:	(Anonymisert)		

Deltaker

Ola Oppegaard, Fredrik Thompson

Informasjon fra deltaker

Tittel *:	Saving Shocks, House Prices and Limited Access to Investment in Local Publicly Traded Firms.
Navn på veileder *:	Plamen Nenov

Inneholder besvarelsen konfidensielt materiale?: Nei

Kan besvarelsen offentliggjøres?: Ja

Gruppe

Gruppenavn: (Anonymisert)

Gruppenummer: 269

Andre medlemmer i gruppen:

BI Norwegian Business School
Oslo, Spring 2022

Saving Shocks, House Prices and Limited Access to Investment in Local Publicly Traded Firms.

Evaluation of the "Sand States" Hypothesis using the Covid-induced Saving Shock

Ola Oppegaard & Fredrik Thompson

Supervisor: Plamen Nenov

Master thesis, Master of Science in Business

Major: Finance

BI NORWEGIAN BUSINESS SCHOOL

Acknowledgements

This thesis marks the culmination of our 5 years of studies at BI Norwegian Business School. In this regard, we would like to thank our families for their support, as well as fellow students and friends for making our time in Oslo memorable in so many ways. We also want to thank BI, and especially the Department of Finance, for a truly inspiring programme of study.

Throughout the process of working on this thesis, we have received invaluable guidance and support from people worthy of acknowledgement. Foremost, we want to thank our supervisor, Plamen Nenov. Without your encouragement and constructive feedback, we would not have been able to write this thesis.

BI Norwegian Business School

Oslo, June 2022

Ola Oppegaard

Fredrik Thompson

Abstract

In this thesis, we examine the relationship between stock market investment and real estate investment, with roots in the *Home bias* theory. We explore in detail, the connection between house price growth in states and areas defined by 3-digit zip-codes in the U.S. and the number of local publicly traded firms. Our thesis uses the Covid-19 pandemic induced saving shock and demonstrates how house prices in areas with a high number of local publicly traded firms tend to rise less compared to house prices in areas with a low number of local publicly traded firms.

In our analysis, we utilise quarterly panel data for both states and areas defined by 3-digit zip-codes, and use a binary- and continuous treatment difference-in-differences model. First, we find a negative relationship between changes in house prices and the number of local publicly traded firms. These findings are in line with existing literature and supports our hypotheses. We also find that the saving shock during the Covid-19 pandemic led to a higher difference in house price growth between zip codes with a high/above median number of headquarters and the low/below headquartered ZIPs. There are mild pre-trends in the data, suggesting our findings may be influenced by pre-existing trends. However, on a general basis we observe a stronger effect after the Covid-19 shock than the pre-trends suggest which supports an increased effect in the savings shock period.

Keywords – Home Bias, Real Estate, US housing cycle, Difference-in-differences, Covid-19, Saving shocks

Contents

List of Figures

List of Tables

1 Introduction

During the Covid-19 pandemic, house prices rose in 54 out of 60 country-specific housing markets which the IMF (2021) published housing statistics on. Particularly in the U.S., Pacific, and some parts of Europe, house prices have continued to rise strongly into today.

The saving shock we saw during the pandemic led to an unprecedented short-run accumulation of household savings. As seen from Figure 1.1, The Personal Saving Rate in the U.S. increased over 6 times from April 2019 to its peak in April 2020. The sharp increase in savings was a result of the uncertainty regarding the pandemic, but also as a response of the fiscal- and monetary stimulus whereby interest rates were set to zero, and there were substantial fiscal transfers to households. Behavioral shifts, and non-pharmacological interventions such as lockdowns was also an assisting factor. The increased savings rate for those who were able to save created an inequality gap, and high-income households accumulated much greater savings during the pandemic. However, governments around the world created aid programs aiming to help those who were not in the high-income group. Bearing in mind some of the consequences of the pandemic, we are interested in where the savings were invested, and what the effect of the increased savings had on asset prices. What is true, is that the Pandemic led to a tale of two economic outcomes. Those who were able to save, and those who struggled to make ends meet (Dossche and Zlatanos, 2020).

So, where did those who were able to save invest their money? The increased saving rate opens up for investment choices. Articles such as (Coval and Moskowitz, 1999) suggest that local bias is an influential factor in where investors tend to invest their money. The phenomenon is present in the capital market where investors and fund managers invest in locally headquartered companies, but also in the real estate market where investment homes tend to be bought close to the investor's primary dwelling. On the next page, we present figure 1.2 which shows how house prices in areas with a low number of local publicly traded headquartered firms have moved compared to areas with many headquarters of public firms in the last 27 years. What is interesting is that, on average, the areas with a low number of local headquarters have risen much more than areas where there are a high number of local headquarters after the pandemic started.

Figure 1.1: FRED - Personal Saving Rate

This graph illustrates the personal saving as a percentage of disposable personal income (DPI), and is retrieved from the Federal Reserve Economic Data. The graph is frequently referred to as "the personal saving rate," and is calculated as the ratio of personal saving to DPI. Personal saving is equal to personal income less personal outlays and personal taxes; it may generally be viewed as the portion of personal income that is used either to provide funds to capital markets or to invest in real assets such as residences. Source: U.S. Bureau of Economic Analysis

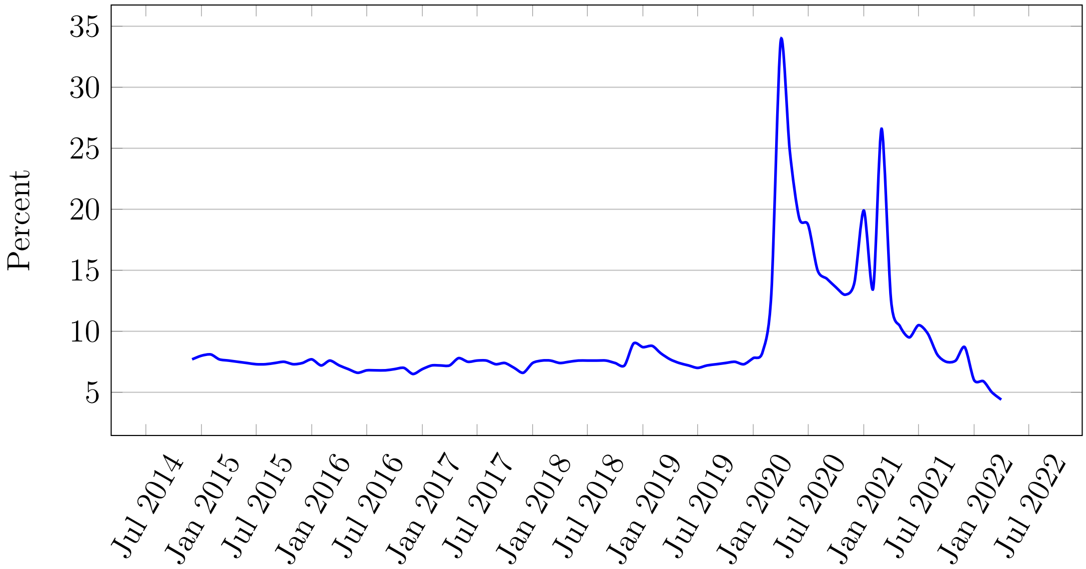
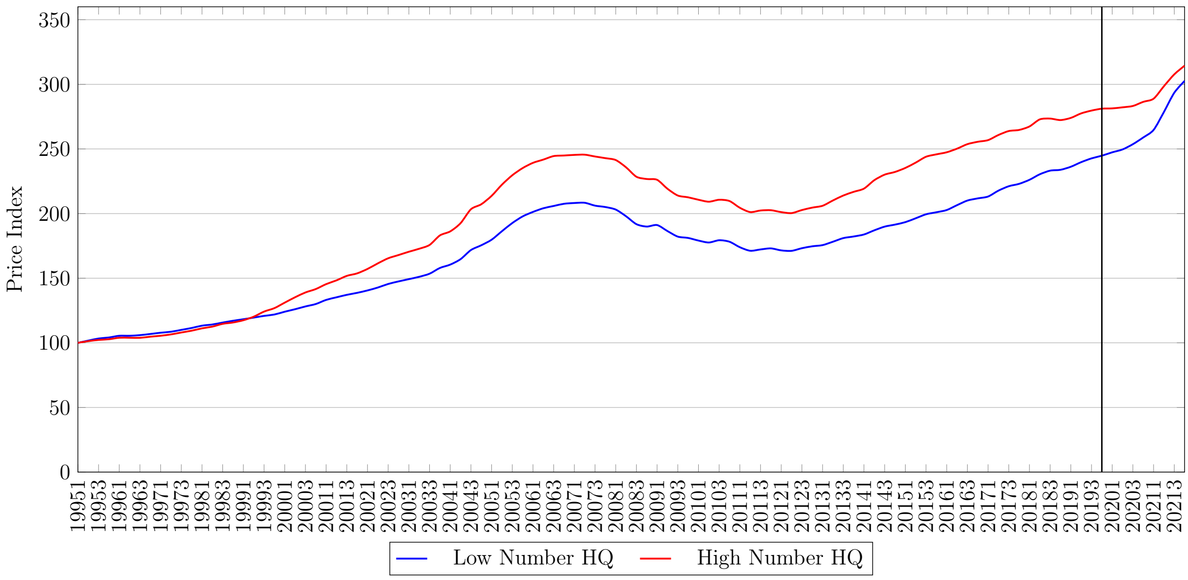


Figure 1.2: Raw HPI Trends

This graph shows the House Price Index (*HPI*) over time for our two defined groups: [1]High number of Headquarters [2]Low number of headquarters. Both groups start out at index 100 in the first quarter of 1995, and ends in the last quarter of 2021. The black line indicates the period *2019 Q4* and hence separates the period before and after Covid-19.



We do not have to travel longer than Sweden to see a difference in saving preference. A research paper conducted in 2017 (Abdli, 2017) showed that down payment of mortgages is the preferred saving method for Norwegians. In fact, 40% of the Norwegians in the survey stood behind the statement. When turning our nose towards Sweden, the view on best-practice changes. Although 30% of the Swedes believe that cash savings is the best option, 28% of the Swedes believe saving in funds is the best option. To compare, only 12% of the Norwegians have funds savings as their preferred saving option. When asked about their actual savings, however, 29% of Norwegians responded they have funds savings, and 14% save in stocks. In Sweden, the actual saving is 40% in funds and 18% in stocks. Part of the motivation behind this thesis is that we want to know what drives the attitude towards savings. Sweden has been an innovative country for decades, with large, publicly traded firms. Norway, on the other hand, have enjoyed the income from the offshore oil sector, and its partly state-owned oil and gas cash-cow Equinor. What is true, is that it seems as if the concentration and the proximity of locally headquartered publicly traded firms have an important effect on whether individuals save in the real estate market or in the capital market.

Choi et al. (2016), referred to as the *"Sand States"* article on the next page, found that households living in Metropolitan Statistical Areas with a low supply of locally headquartered publicly traded firms tend to invest in the local housing market. That article is the bridge between the findings between Norway and Sweden, and the attitude towards investment choices in the United States. We base our hypothesis upon that article, as well as articles on stock participation, local bias and financial literacy. We believe the area of study is particularly interesting because it is counterintuitive to what we generally perceive as triggers in the real estate market. A counterweight to the drivers in the U.S. is our own hometown, Oslo. The capital of Norway has a large number of headquartered companies, as well as a strong real estate market that appreciated profoundly before, during, and after Covid. Additionally, we have seen that smaller places are reliant on cornerstone companies to have growing real estate markets, especially cities with cyclical industries that employ a major part of the city's work force and contribute greatly to the total GDP of the municipality. In 2015, the wealthy city of Stavanger was hit by falling oil prices which directly hit the real estate market. As of today, Stavanger is still recovering from the crisis, and its five-year growth (2015-2019) is -2.9% according to statistics from (SSB, 2022), which is extremely low compared to Oslo's 39.7% rise in the same period.

We split our research in two parts. Firstly, we want to figure out if the findings from the sand states article also applies to 3-digit ZIP codes, and not only MSAs across the US. We also want to figure out if the Covid saving shock led to an increase in house price growth for the below median/low headquartered ZIPs versus the above median/high headquartered ZIPs.

Specifically, our hypotheses are stated as follows:

H1: ***"Does differential number of locally headquartered publicly traded firms across ZIP codes have a significant effect on the differences in house price growth in the U.S?"***

and

H2: ***"Has the saving shock of Covid-19 led to increased differences in house prices in areas with a low number of locally headquartered publicly traded firms against areas with a high number of locally headquartered publicly traded firms?"***

We structure the thesis in the following way. In the next chapter, we will provide background information on home bias, and the ways in which it could influence markets by reviewing the existing literature on the topic. The third chapter contains a description of the data, variables, and methodology used in the analyses. In the fourth chapter, we present and discuss the findings of our analyses. The fifth chapter contains our conclusions, as well as potential limitations with our analyses and suggestions for future research.

2 Literature Review

In the following chapter, we provide the relevant background for this thesis. What is true from reading the existing literature is that there are a lot of determinants that make up an investor's saving profile. The local bias/ home bias is an important factor in determining the distance between the investor and the investment object, regardless of the investment. In regards to what type of investment one chooses, it seems that both literacy and availability are two important factors. We want to use this knowledge and check if these trends also apply at the ZIP-code level in the U.S, as there are different requirements in regards to urbanity and population for MSAs and ZIP-codes. We also want to figure out if the patterns we have seen in normal times gets amplified when a saving shock occurs, as with the Covid-19 pandemic.

The housing market in the U.S. has been a significant factor during world crises, and the housing cycle in the 2000s was one of the leading drivers of the financial crisis in 2008. Choi et al. (2016) discusses how the excesses were concentrated in Arizona, California, Florida and Nevada, known as Sand States due to their abundance of beaches or deserts. Three main points have been listed as explanations for the housing cycles in the MSAs. Firstly, the easing of lending standards, particularly in low-income growth areas (Mian and Sufi, 2009). The second discusses the zoning issues in the sand states that have created an inelastic supply (which can amplify the volatility of housing prices). The third discusses that the house price increases were over-extrapolated.

However, these three possible explanations cannot justify the phenomenon in the sand states. For instance, MSAs with low income growth during the 1990s and MSAs with low housing supply elasticity had bigger housing cycling excesses. In a multiple regression model, however, these factors do not explain much of the Sand State effect. Other factors such as population growth in the areas also cannot explain the excesses that were seen. A key fact worth mentioning is that there tends to be a local bias in households' investment, also with regards to the capital market. Households do not tend to diversify their portfolio. A study conducted by French and Poterba (1991) found that households tend to hold portfolios of stocks headquartered less than 60 miles of where they live. The bias towards proximity investment is even stronger in buying investment homes. Choi et al. (2016) demonstrated that the median distance between an investor's primary residence and their investment home is 10 miles, which is a lot less than the 220 mile median distance between a vacation home and a primary residence.

Choi et al. (2016) proposes an explanation of the Sand States effect involving the local investment opportunity of investors. More specifically, the paper studies whether the availability of publicly traded firms in an MSA, measured as the book value of firms headquartered in an MSA relative to total income of that MSA matters for house prices. On average, Sand States have low residual book value, which implies a low supply of equities relative to total income. From this, the hypothesis is that MSAs with few locally headquartered stocks are more likely to invest in investment homes. The article proposes an only-game-in-town effect which effectively is a result of the local bias. Naturally, the residents of these MSAs have the option of buying national stocks, but have resistance towards doing so.

There are some reasons why the only-game-in-town effect occurs. Firstly, the lack of publicly traded locally headquartered companies can drive their stock price up, making them an expensive option which causes a shift towards local real estate. The next possible explanation is that people living in MSAs with few publicly traded firms may not have the same financial background. What makes the subject even more interesting is that investment homes have experienced an increased portion of the housing market in recent years. From 2003 to 2005, primary listings' share of the total market dropped from 67% to 60%, whereas investment homes rose from 22% to 28% (Choi et al., 2016), giving more room for speculation in the total market. By using the residual book value, the analysis show that the variable can account for a large portion of the house price excesses in the Sand States. The analysis also found that it can account for 11% of the mortgage defaults and excess home price fluctuations.

The regression consisted of 277 MSAs, and is population weighted. In the MSAs with a population equal to or over 750 000, the effect of the residual book value is stronger and can explain all the excesses in mortgage origination, defaults, and price fluctuations.

By reviewing demographic data, the authors of the paper find that households living in MSAs where there are few locally public traded firms headquartered buy more investment homes and less stocks compared to other households. What is a concern regarding this strategy is that MSAs with few local firms might have more low FICO type households or non-white households who prefer homes to stocks. When controlling for the potential bias, they find that even in high FICO households, investments are shifted towards real estate in MSAs with few local firms which is in line with the hypothesis and the only-game-in-town effect.

Coval and Moskowitz (1999) show how preferences for investing close to home also applies to portfolios of domestic stocks, and how investment managers prefer firms with local headquarters. There are a few reasons why investors generally prefer to invest locally. Factors like information bias are important, because investors believe they are better informed regarding local companies rather than international ones. Brokers also tend to recommend investing in local companies since they potentially have a relation to local management, hence benefiting from keeping money locally. There are two governing factors behind the theory of preference for local investments. Firstly, there are those who rely on national/governmental frictions. Secondly, there are those who rely on frictions associated with distance.

Based on the results in this paper, geographical distance seems to play an important role in a fund manager's portfolio choice. There are also important characteristics that local firms often have. To generalize, local firms tend to be nontraded-goods producing firms, with high leverage. Worth mentioning, however, is that there are some pitfalls one should avoid when addressing the effect of distance. First of all, a New York-based investor can favor local stocks unrelated to geographical proximity. The choice can be based merely on the fact that they are situated in the finance capital of the world. Secondly, there is a big difference between economical distance and geographical distance. Los Angeles to New York is farther in distance than Los Angeles to El Paso, but the economical distance is a lot closer. However, the study's findings on proximity in relation to investor portfolio choice are important and open up for further research and discussions.

Even though international diversification is globally recognized, investors tend to hold their wealth in domestic assets. An important aspect when addressing whether to diversify globally or not is to investigate the expected returns across nations. French and Poterba (1991) find that a British investor must expect annual returns in the U.K. more than 500 basis points above those in the U.S. market to justify their 82% investment in domestic shares. The reason is the higher standard deviation in the British market compared to the American market. For an American investor, the expected annual returns must be 250 basis points above the expected returns on Japanese stocks. On the flip side, the return for a Japanese investor must be 350 basis points above the expected return on U.S. stock.

Another interesting study was conducted, where the authors wanted to figure out at what percentage of expected return an investor would hold an equally weighted international value portfolio. These results also suggest that investors expect higher domestic returns rather than returns coming from a diversified international portfolio. The reasons why investors tend to overweight domestic markets have two main explanations according to the paper by French and Poterba (1991). Firstly, institutional factors may reduce returns from investing abroad or they may limit investors from investing abroad. Secondly, investor behavior for i.e. expected return varies systematically across different investor groups. Another aspect is the perception of risk in the equity market. Investors may take on extra risk in the foreign market because they know less about foreign markets and firms. Lastly, the paper relates the behavior to the housing market with the example that people tend to own residential real estate where they work, which is a striking example of incomplete diversification and far worse than holding a national real estate portfolio. (French and Poterba, 1991).

Lewis (1999) discusses what she calls "consumption home bias" and "equity home bias". The article finds that investors who steer their equity holdings away from foreign assets will not optimally diversify their home output risk. When trying to figure out the low degree of risk-sharing, there are a few governing explanations that seem plausible. Firstly, the absence of non-tradables, since its presence is exclusively in the domestic country. Another plausible explanation is that the risk-sharing gains do not surpass the costs of diversification. Research has been conflicting on the matter, but the latest research has shown that the gains from diversification exceed the costs. Furthermore, the article discusses home bias in equities and consumption. The theory of home bias in equities says that investors hold too few foreign assets relative to an optimal portfolio that would hedge risk and even possibly increase returns. Home bias suggests that output risks are not optimally shared across nations, and domestic consumption is correlated with country-specific shocks. The basic intuition would suggest that these two puzzles are linked, but home bias in equities is neither sufficient nor necessary for home bias in consumption.

As a result of the financial innovations in recent years, the portfolio choices have expanded. Iachan et al. (2021) investigate what effect this increased opportunity has led to. The result suggests that an investor who gains access to greater portfolio choices increases his savings because greater choice enables earning of aggregate risk premium or undertaking of speculative position, leading to greater risk-adjusted returns. The article also finds that the increased speculation that comes as a result of customization opportunities leads to more dispersed portfolio returns. Furthermore, the paper suggests that providing households with a greater choice of portfolios could improve welfare by countering frictions that lower savings. At the same time, households with heterogeneous beliefs use greater choice to load onto non-systematic risk, as quoted from the article. The result is that the risk in consumption reduces welfare which implies a net effect of financial innovation on household welfare that is likely to be ambiguous.

Polkovnichenko (2004) studies the effect limited stock market participation has on equity premium. While existing literature suggests that limited participation can yield a high equity premium, this article suggests the opposite. It does so by showing how the size of the equity premium in a limited participation economy depends on the assumptions about the structure of non-financial endowments. The model used in the paper includes appropriately calibrated labor income of stock market participants along with income from dividends. He found that aggregate labor income is weakly correlated with aggregate dividends as the result of the consumption of shareholders being not volatile, nor has a strong correlation with dividend growth to generate a high equity premium.

The introduction of new financial products and services aims to increase stock market participation. However, some of these products are hard to grasp - especially for those with little financial literacy. At the same time, the decision-making responsibility is currently shifting towards private individuals and away from governments and employers. Van Rooij et al. (2011) find that a lacking understanding of finance and economics significantly decreases stock ownership. The consequence of missing out on the stock market can lead to a welfare loss. On the other hand, it is not given that unsophisticated investors can take full advantage of the stock market if they would be involved, with a low understanding of i.e. diversification. What is clear, however, is that improved financial literacy is important for people to take ownership of their pensions and savings, when individual responsibility becomes more important than before.

Bearing in mind that households can invest in three asset classes: stocks, bonds and real estate; we know that only a subset of investors can invest in the stock market. Rieger (2017) studied a model with limited market participation with heterogeneous beliefs, and two consumption goods. What the model showed was that with full participation, the volatility of real estate and stocks was reduced. The paper also studied how relaxing credit constraints affects volatility. What the author found was that the price/rent ratio always increases with the relaxed credit constraints. If credit constraints are loose enough, volatility under full participation is increased. Thus, indicating that a sharp increase in house prices was not only due to relaxed credit standards but also increased financial market participation.

3 Data and Methodology

In this section, we present both the data and the methodology used in our analysis. First, we elaborate on the sources of our different data and the structure and size of the samples derived from these sources. Second, we present the variables and elaborate on the specific methodology that we apply in our analysis. For the continuous treatment DID model, we want to investigate the movements in the housing market between areas with a high density of locally headquartered publicly traded firms versus the areas with a low density of locally headquartered publicly traded firms. We also conduct this analysis on the binary treatment DID model, determining ZIP-codes above and below the median number of headquarters. Thereafter, we analyze the effect that the Covid shock had on these areas.

3.1 Data

We have chosen to collect Housing Market Indices (HPI) on a quarterly frequency for three-digit zip codes, and then match the number of headquarters for each three-digit zip code. Each zip code is also connected to its corresponding state, making it possible to include state fixed effects. As shown by Choi et al. (2016) we know that the asset value is closely correlated with the total GDP of an MSA. We therefore want to explore the effect of adding *IPC* - Income Per Capita as a control variable. In the following three subsections we elaborate on the issues and benefits of our choices on data collection for *HPI*, Zip codes and *IPC*. To start off, we present a summarizing table of our variables in table 3.1 below.

Table 3.1: Descriptive Statistics

This table provides the Number of Observations, Median, Mean, 95% and 5% percentile, Standard Deviation, Variance, Skewness and Kurtosis for our data on *HPI* - House Price Index, *HQ* - Headquarters and *IPC* - Income Per Capita from the first quarter of 1995 until the last quarter of 2021 in our 409 different three digit Zip-codes across the U.S. *HPI* is the dependent variable and *HQ* and *IPC* are the independent variables.

	HPI	**HQ**	**IPC**
Nr. Obs	44,064	409	929
Median	165.62	5	14,152.83
Mean	177.8	20.01	15,905.2
5% Percentile	104.39	2	2,719.37
95% Percentile	303.03	76	36,226.41
Std. Dev	63.38	61.02	10,819.04
Skewness	1.45	8.93	0.98
Kurtosis	5.89	106.08	4.25

3.1.1 House Price Index (HPI)

In our thesis we have chosen to use data on house prices from the Federal Housing Finance Agency (FHFA, 2022). FHFA publishes house price indices that measure changes in single-family home values based on data from all 50 states that extend back to the mid-1970s. The FHFA HPI analyzes tens of millions of home transactions and provides information on price changes at the national, census division, state, metro area, county, ZIP code, and census tract levels. To assess house price transaction data, the FHFA uses a transparent methodology based on a modified version of the weighted-repeat sales (WRS) methodology proposed by Case and Shiller (1990). According to their paper, the WRS method is based on the assumption that the logarithmic price P_i, of the ith house at time t can be viewed to have three components:

$$\mathrm{P}_{i,t} = \mathrm{C_t} + \mathrm{H_{i,t}} + \mathrm{N_{i,t}} \tag{3.1}$$

Where:

- C is the log of the citywide level of housing prices at a particular time, which captures the location or market effect on price.
- H captures the impact of individual property-related factors on price and is assumed to follow the Gaussian random walk.
- N is an identically distributed normal noise term (which has zero mean and variance σ_N^2).
- It is further assumed that C, H, N are uncorrelated.

FHFA delivers more information than other house price indices due to the size of the sample. One of the most well-known alternatives is the SP CoreLogic Case-Shiller Home Price indices. Both indices use distinct data and measuring approaches, resulting in varied outcomes. FHFA, for example, gives equal weight to all homes, whereas the SP CoreLogic Case-Shiller Home Price indices are value-weighted (SP, 2022). Furthermore, unlike the Case-Shiller index which only considers purchase prices, the HPI incorporates refinance appraisals as well making the HPI offering more coverage.

3.1.2 ZIP codes and Headquarters

There are approximately 44,000 five-digit ZIP codes and 929 three-digit zip codes in use in the U.S. The data on headquarters per ZIP code is retrieved from Compustat. Compustat is proven to be a reliable source, and their errors dropped markedly after 1970 (Bennin, 1980). The number of Tickers then represents the number of Headquarters in each area.

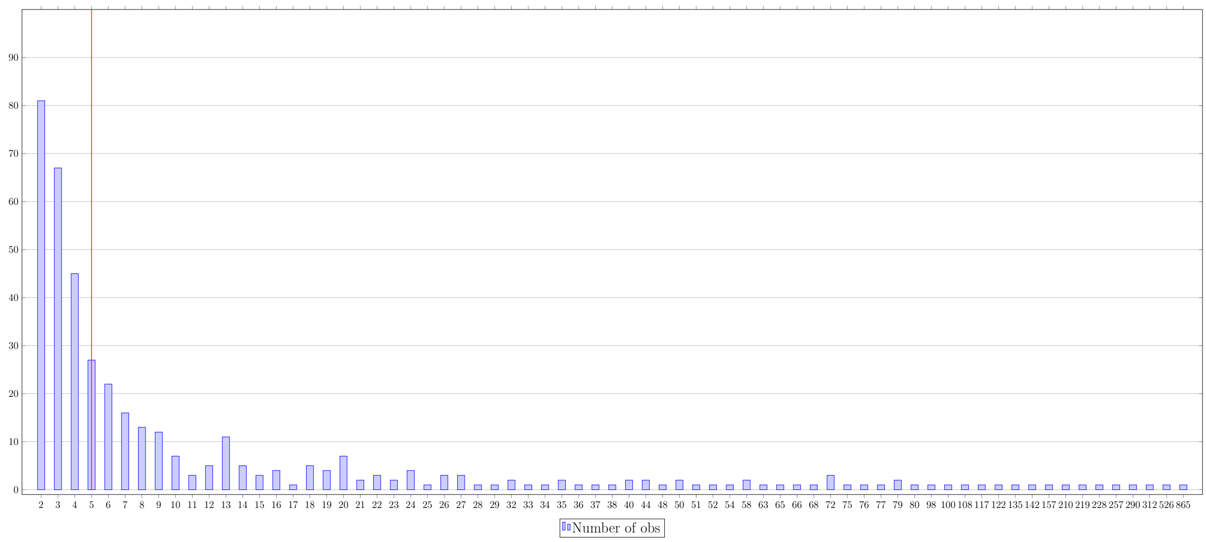
Zip codes are not randomly distributed across the U.S., and it is important to note that there are large differences between these areas when comparing sizes. The average land size of the U.S. per ZIP code is around 82.25 square miles for five-digit ZIP codes. The tiniest ZIP code is 0.1 square mile and includes just two blocks in Long Island City. On the other hand, ZIP code 89049 has the largest land area, covering 10,000 square miles in Nevada. The main reason for these differences is that ZIP codes are mainly based on people rather than topography. We can divide our data on zip codes the following way:

$$\underbrace{\boxed{1}}_{\text{National Area}}\ \underbrace{\boxed{2}\ \boxed{3}}_{\text{Regional Area}}\ \underbrace{\boxed{4}\ \boxed{5}}_{\text{Delivery Area}}$$

The first digit is the "National Area", and the U.S. ZIP codes have ten different national areas. The next two digits are "Regional Areas", and represent smaller areas within states. There are on average ten "Regional Areas" per state. For a more detailed outlook, the zip codes within the states are illustrated in appendix A3. Since zip codes are artificial, and were created to make mail distribution easier, there was a conscious attempt to make them population-wise uniform. This helps the three-digit zip code to not have abnormally large tails on population sizes. There is still a great range between the number of headquarters between zip codes, with New York at the top with zip code *100* that has 865 headquarters to Wyoming at the bottom with zip code *820* that has 1 headquarter. We do not have headquarter observations in all 3-digit ZIP codes, which is due to ZIP codes without publicly traded headquarters in its area. Because of this we have matching data for all variables on 409 unique zip codes. To better visualize the range of registered observations on headquarters in each three digit zip code, a histogram is presented in Figure 3.1.

Figure 3.1: Histogram of observations on headquarters

This histogram illustrates the amount of headquarters that occurs within a specific amount. The range of outcomes are presented along the x-axis. The y-axis represents the number count of occurrences on headquarters. The red line represents the median value.



3.1.3 Income Per Capita

Income per capita is the mean income computed for every man, woman, and child in a particular group including those living in group quarters. It is derived by dividing the aggregate income of a particular group by the total population in that group. This measure is rounded to the nearest whole dollar. The data retrieved on *IPC* is gathered from the United States Census Bureau which conducts the *American Community Survey* generating per capita income for 2019 (CensusBureau, 2019). We use the variable named *S1902C03019E* in the Bureau's dataset which is the mean per capita real income for five digit zip codes. By taking the weighted sum of mean *IPC* and matching it with its corresponding 3-digit zip code we then have the mean per capita income at the 3-digit zip codes.

Some studies have found that the American Community Survey (ACS) have less precise income estimates at central city neighborhoods, and that there are very poor or very rich neighborhoods that have lower quality income estimates according to Spielman et al. (2014). Bearing in mind that there could be some flaws in the data, statistics from all surveys are subject to sampling and non-sampling errors. We find *IPC* for all 929

3-digit zip codes, but due to Compustat's lack of information on headquarters described in chapter 3.1.2 we are only able to have full information on 409 unique zip codes.

3.2 Methodology

We will use a difference-in-differences model that compares the evolution of house prices in locations that have a smaller number of headquarters of publicly-traded firms compared to locations with a higher number of headquarters of publicly-traded firms before and after the first COVID-19 wave in the U.S.

By studying this effect, our model tries to calculate the effect of a saving shock by comparing the average change over time in the outcome variable for areas with a low degree of local publicly traded firms to the average change over time for areas with a high degree of local publicly traded firms. We have constructed both a binary- and continuous treatment difference-in-differences model. Our "shock" is set to be in the first quarter of 2020. The Covid-19 pandemic does not simply turn "on" and has varying intensity. A continuous model may therefore offer practical advantages compared to a binary model in that variance in shock-intensity allows for the evaluation of shocks.

The regressions are conducted using quarterly house price index data. We create D_z as a binary variable that equals 1 if the number of local publicly headquartered firms is above the median value across all zip codes in 2019 and 0 otherwise. These different states for D_z are referred to as $Zip_{i,Above}$ and $Zip_{i,Below}$ for the binary regression. We estimate the binary treatment DID model and regress from one quarter all the way up to eight quarters ahead, as Covid was a long lasting Pandemic. To check the pre-trends, we follow the same methodology but create lagged variables instead of forward variables. Last, we estimate the continuous treatment model by changing the dummy variable D_z to ln_{HQ}, and holding all else equal.

3.2.1 Binary Treatment DID Model

$$\Delta^h \ln p_z = \alpha_h + \beta_h D_z + \gamma_h IPC_z + \delta_{s(z)} + \epsilon_{zh} \tag{3.2}$$

where $\Delta^h \ln p_z$ is the h-quarter log change of the house price index relative to the fourth quarter of 2019 for zip z. D_z is a binary variable that equals 1 if the number of local publicly headquartered firms is above the median value across all zip codes in the fourth quarter of 2019 and 0 otherwise. IPC$_z$ is the Income Per Capita for 2019 for zip z. $\delta_{s(z)}$ denotes the state fixed effects. α_h is the intercept for horizon h. Note that α_h also absorbs the common time trend shared by both groups, i.e it absorbs the nation-wide trend in the house prices. Finally, ϵ_{zh} is the residual term for $h \in \{1, 2, 3, ..., 8\}$

3.2.2 Continuous Treatment DID Model

$$\Delta^h \ln p_z = \alpha_h + \beta_h \left(ln(HQ_z)\right) + \gamma_h IPC_z + \gamma_{s(z)} + \epsilon_{zh} \tag{3.3}$$

Where $\Delta^h \ln p_z$ is the h-quarter log change of the house price index relative to the fourth quarter of 2019 for zip z. $ln(HQ_z)$ is the logarithm of the number of locally headquartered firms. IPC$_z$ is the Income Per Capita for 2019 for zip z. α_h is the intercept for horizon h. ϵ_{zh} is the residual term for $h \in \{1, 2, 3, ..., 8\}$

3.2.3 Assumptions

DID is subject to all of the OLS model's assumptions. DID also necessitates the assumption of parallel trends. According to the parallel trends assumption, the difference between the areas with few publicly traded firms and many publicly traded firms must remain constant throughout time in the absence of a shock. The parallel trends assumption says that although the price level can be different, the house price trend has to be the same absent of a shock. So if one area's house prices change by x%, the control area also needs to change by x% absent of a shock. The areas do not have to overlap, but they have to move together. There is no statistical test for this assumption, and it is therefore vital with a thorough visual inspection of the plots. It has also been suggested that the shorter the time period under examination, the more likely the assumption is to hold according to Rambachan and Roth (2019). If we observe the same trend for both the "control-area" and the "treatment-area" under a shock it could be reasonable to assume that local publicly traded firms do not have any effect. Likewise would a difference in the trend assume that there is an effect, but a possible violation of the parallel trend assumption could lead to biased estimation of the causal effect.

To get conclusive estimates, the makeup of areas must remain as unchanged as possible throughout time to ensure the accuracy of the DID estimate. A possible threat that may affect our results could be caused by autocorrelation and/or Ashenfelter dips. The last means that if some area's house prices drop before the shock the results would be biased. Likewise, the results will be biased if the prices deviate greatly before the shock. To take this into account the pre-trends are presented in chapter 4.2.1.

4 Analysis

In this chapter, we present the results of our regressions. Second, we present a robustness check and model our pre-trends before we end with a discussion of the limitations.

4.1 Main results

In this section, we present our regression results. We start by looking at the raw trends in Figure 4.1 and 4.2. These graphs are an extended analysis of Figure 1.2 presented in the introduction, and illustrates how the house prices grew differently for our two groups. To then examine our hypotheses in more detail, we present the results from our DID model. The binary treatment DID model is presented in chapter 4.1.1 where we do not control for *IPC*. In the next section - 4.1.2 we use *IPC* as a control variable in our binary treatment DID model. In sections 4.1.3 and 4.1.4 we do the same for our continuous treatment DID model and present the results both with and without controlling for *IPC*. Since we are worried about the variations across states, we want to include state fixed effects in our regression to control for state specific trends. We believe this is especially important in the US. because of the profound gaps in economic and demographic conditions across states. Ultimately, what we are left with is looking at the within state variation and getting rid of between state variation, meaning that we compare zip-code deviations from the state's own mean. Lastly, we give a summary of our findings in section 4.1.5.

Figure 4.1: Raw Trends

This graph shows the *HPI* over time for our two defined groups: [1]High number of Headquarters [2]Low number of headquarters. Both groups start at different index levels at fourth quarter of 2017, and ends in the last quarter of 2021. The black line indicates the period *2019 Q4*. The dotted line illustrates the predicted trend of group [1] without the impact of Covid-19.

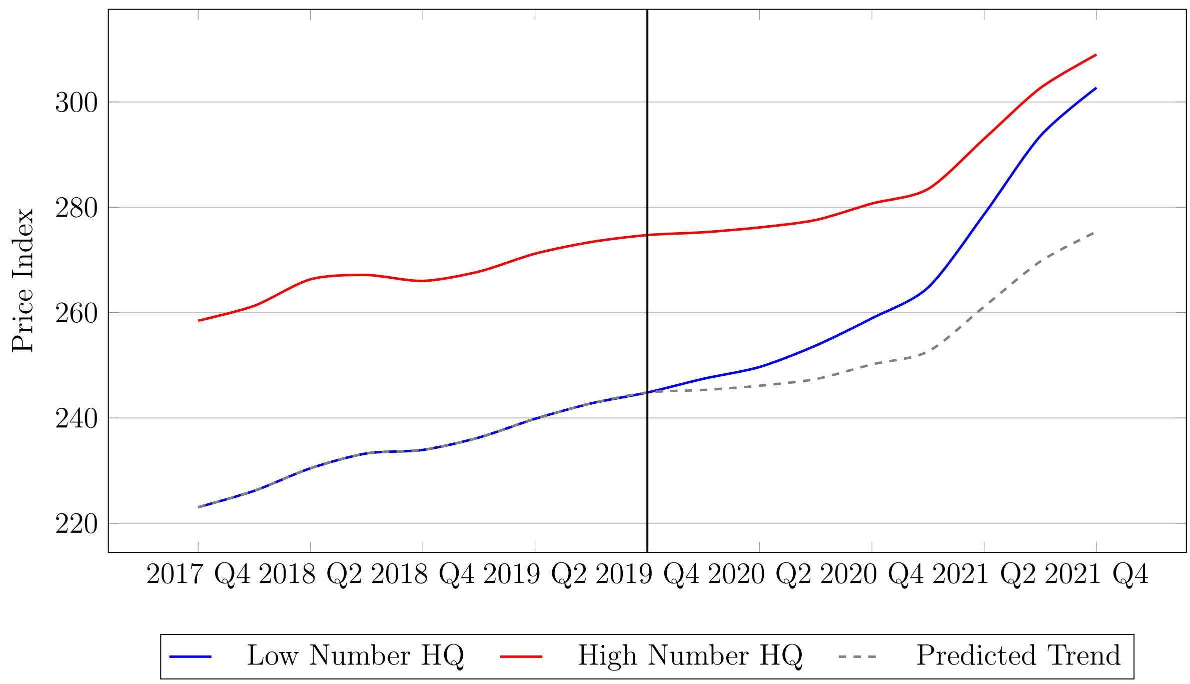
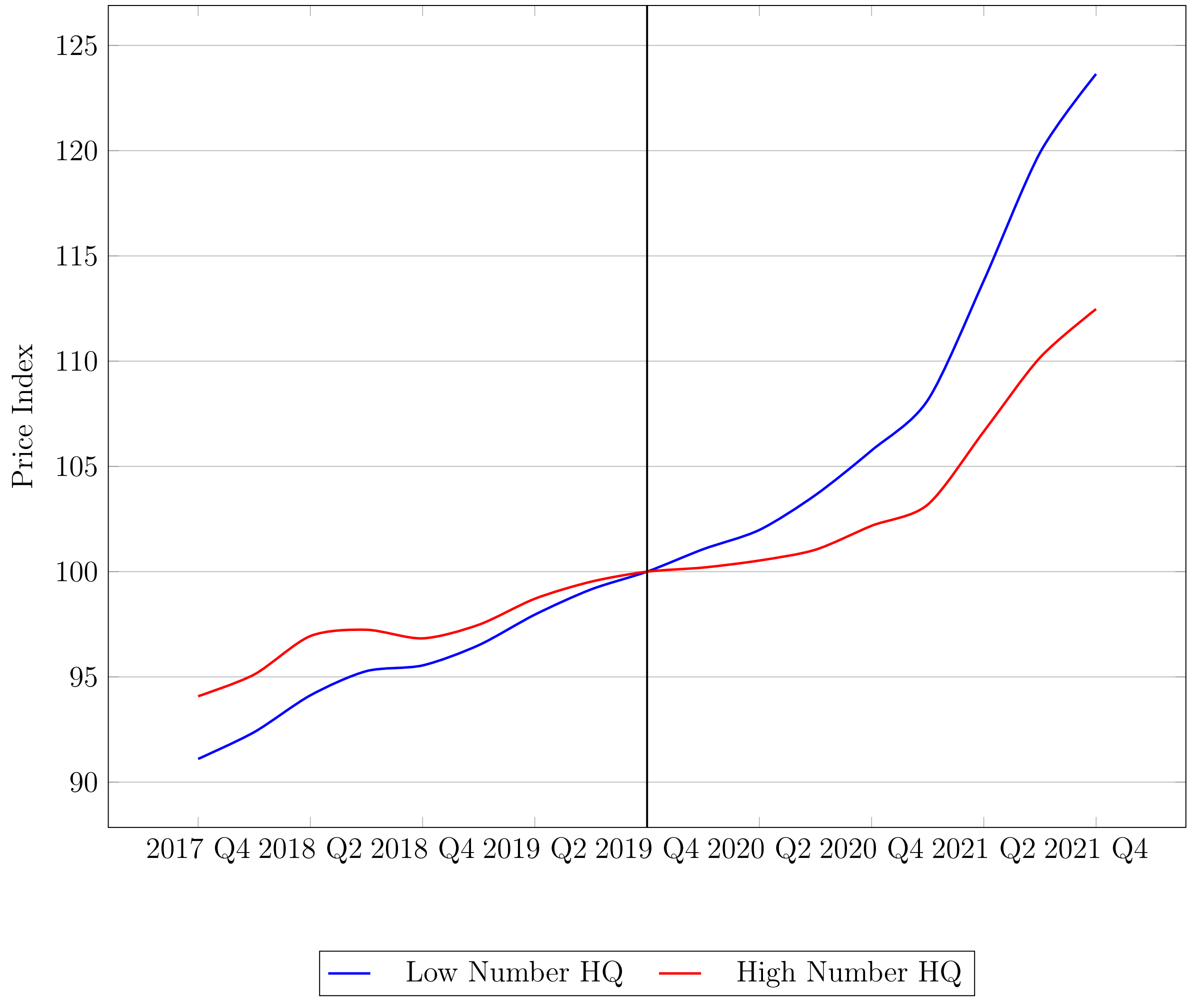


Figure 4.2: Raw Index-adjusted Trends

This graph shows the *HPI* over time for our two defined groups: [1]High number of Headquarters [2]Low number of headquarters. Both groups have index 100 in the fourth quarter of 2019 which is indicated with the black line.



4.1.1 Binary Treatment DID Model

The results from estimating equation (3.2.1) for $h \in \{1, 2, 3, ..., 8\}$ are presented in table 4.4. The results range from -0.032 for $h = 1$ to -1.327 for $h = 8$ which represents 2021 Q4. Since the results deviate from 0, this indicates that there is a difference between the post-pandemic growth in house prices between $Zip_{i,Below}$ and house prices in $Zip_{i,Above}$. We notice that we get significant results in the last quarter of 2020, as well as the first, third and fourth quarter of 2021. To get a better grasp of the change in $\hat{\beta}_h$, the value of $\hat{\beta}_h$ with its corresponding confidence intervals are plotted in Figure 4.6. Here one can see the trend that $\hat{\beta}_h$ increasingly deviates from 0. One explanation could be that there is a delay effect of the hold-up period in the real estate market.

Table 4.1: Binary Treatment DID without IPC

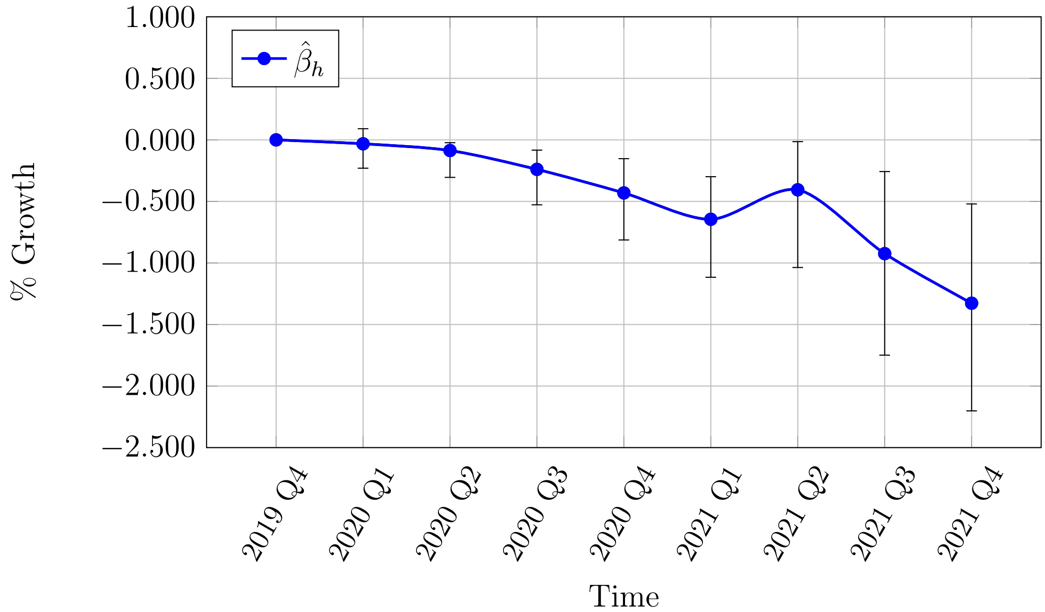
This table provides the coefficient estimates $\hat{\beta}_h$, t-value, F-test, 95% confidence interval - *CI*, R^2 and Adjusted R^2 for the regression done on the quarterly numbers for the amount of headquarters against the house price index in each area. T-statistics are based on robust standard errors. $\hat{\beta}_h$ and its corresponding *CI* are approximated in percent. $\hat{\beta}_h$ represents the difference in house price growth, the greater it deviates from 0, the greater are the differences in house price growth. A negative $\hat{\beta}_h$ indicates that house prices in areas with above median number of headquarters tend to rise less than house prices in areas with below median headquarters after the fourth quarter of 2019.

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4
$\hat{\beta}_h$	-0.032	-0.087	-0.240	-0.431	-0.645	-0.406	-0.924	-1.327
t-value	-0.32	-0.78	-1.64	-2.22**	-2.7***	-1.26	-2.2**	-2.98***
F Statistic	0.1	0.61	2.68***	4.92***	7.26***	1.6	4.85***	8.9***
DF	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)
CI	[-0.23:0.166]	[-0.305:0.132]	[-0.527:0.048]	[-0.813:-0.049]	[-1.116:-0.174]	[-1.037:0.225]	[-1.749:-0.099]	[-2.202:-0.452]
R^2	0.192	0.304	0.438	0.466	0.513	0.557	0.561	0.612
Adj R^2	0.090	0.216	0.367	0.399	0.451	0.501	0.505	0.563

Note: *p<0.1; **p<0.05; ***p<0.01

Figure 4.3: Binary Treatment DID

This plot shows the coefficient estimates $\hat{\beta}_h$ per quarter from 2019 Q4 to 2021 Q4. The error bars are 95% confidence intervals based on robust standard errors. The more $\hat{\beta}_h$ deviates from 0, the higher are the differences between house price changes in areas with above median amount of headquarters ($Zip_{i,Above}$) against areas with below median amount of headquarters ($Zip_{i,Below}$). A negative $\hat{\beta}_h$ indicates that house prices in areas with above median number of headquarters tends to rise less than house prices in areas with below median headquarters after the fourth quarter of 2019.



4.1.2 Binary Treatment DID model with *IPC* as a control variable

From the *"sand states"* article by Choi et al. (2016) we know that the asset value is closely correlated to the MSAs total GDP. We use this to check whether we can control for the inequality gap in the U.S. by adding *IPC* as a control variable. The other reason why we want to control for income is that the pandemic led to two economical conditions: those who were able to save, and those who struggled to make ends meet. Since we know that GDP and asset value are closely correlated, we believe there could be ZIPs that have a population unable to save as their state's asset value suggests a low state GDP.

When regressing the number of headquarters on *IPC*, we get positive results significant from zero, with a t-value of 35.63 and a coefficient of 0.0015. Therefore, for every dollar increase in *IPC*, the number of headquarters per zip code increases by 0.0015. The correlation between the number of headquarters and the *IPC* is 27.5%. The most important takeaway from this is that the relationship between income and headquarters is positive, and hence we are interested in controlling for income. In table 4.2 we present the results and the corresponding statistics for the regression controlled for IPC.

Table 4.2: Binary Treatment DID With IPC

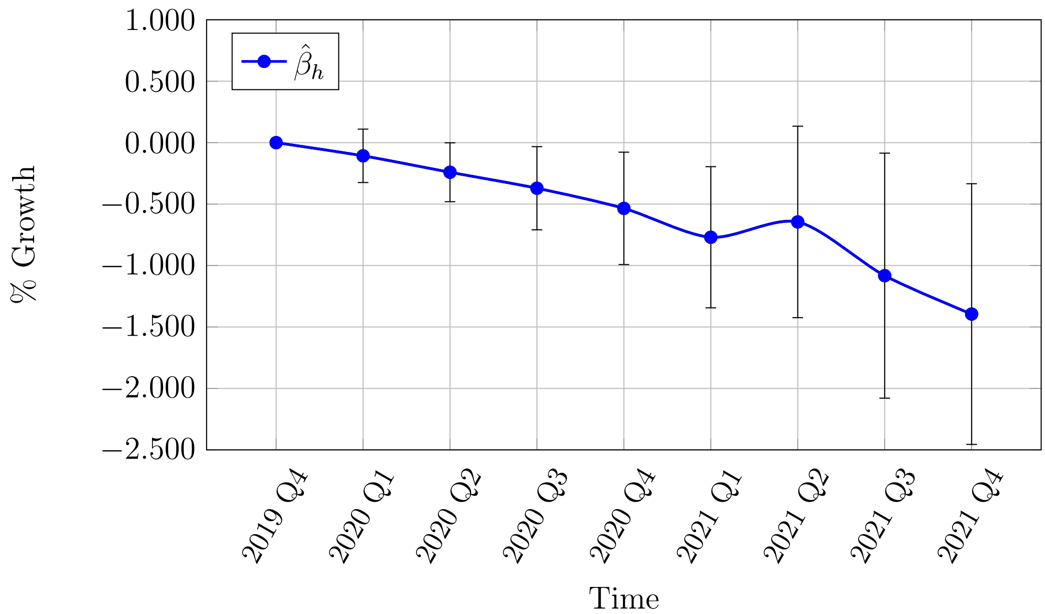
This table provides the coefficient estimates $\hat{\beta}_h$, t-value, F-test, 95% confidence interval - *CI*, R^2 and Adjusted R^2 for the regression done on the quarterly numbers for the amount of headquarters against the house price index in each area. T-statistics are based on robust standard errors. $\hat{\beta}_h$ and its corresponding *CI* are approximated in percent. $\hat{\beta}_h$ represents the difference in house price growth, the greater it deviates from 0, the greater are the differences in house price growth. A negative $\hat{\beta}_h$ indicates that house prices in areas with above median number of headquarters tend to rise less than house prices in areas with below median headquarters after the fourth quarter of 2019.

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4
$\hat{\beta}_h$	-0.107	-0.241	-0.371	-0.535	-0.770	-0.645	-1.082	-1.395
t-value	-0.97	-1.98**	-2.16**	-2.3**	-2.64***	-1.63	-2.14**	-2.59***
F Statistic	1.13	3.79***	2.38**	2.79***	3.89***	1.33	2.55**	4.45***
DF	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)
CI	[-0.325:0.11]	[-0.481:-0.001]	[-0.71:-0.032]	[-0.992:-0.077]	[-1.344:-0.195]	[-1.424:0.134]	[-2.079:-0.085]	[-2.456:-0.334]
R^2	0.198	0.321	0.443	0.468	0.514	0.560	0.561	0.612
Adj R^2	0.095	0.233	0.372	0.400	0.452	0.503	0.505	0.562

Note: *p<0.1; **p<0.05; ***p<0.01

Figure 4.4: Binary Treatment DID with control for IPC

This plot shows the coefficient estimate $\hat{\beta}_h$ per quarter from 2019 Q4 to 2021 Q4. The error bars are 95% confidence intervals based on robust standard errors. A lower $\hat{\beta}_h$ value represents a bigger gap between house prices in areas with above median amount of headquarters ($Zip_{i,Above}$) against areas with below median amount of headquarters ($Zip_{i,Below}$). A negative $\hat{\beta}_h$ indicates that house prices in areas with above median number of headquarters tend to rise less than house prices in areas with below median headquarters after the fourth quarter of 2019.



4.1.3 Continuous Treatment DID model

We want to dig further by changing the binary variable $Zip_{i,Above}$ to $ln(_{HQ})$ which is the logarithm of the number of headquarters per ZIP. The reason why is that continuous treatment DID regression tends to provide more statistical power and are easier to interpret. Results are presented in Table 4.4 and in Figure 4.5 below. We notice that there are results significantly different from zero in most regression intervals suggesting a relatively higher house price growth in the ZIPs with a low number of headquarters compared to places with a high number of local headquarters. This indicates a negative relationship between number of headquarters and the house price index - in line with the results from the binary treatment DID regression, it seems to be a delayed effect as the coefficients increase. This is better displayed by plotting the values of $\hat{\beta}_h$ in figure 4.5. The different values of $\hat{\beta}_h$ plotted tells us how a percentage decrease in HQ will lead to a percentage higher growth in house prices. For example: if $\hat{\beta}_h$ equals -0.5, a 1% decrease in HQ leads to a 0.5% higher growth in house prices, meaning that if you were to double the amount of HQ this will, ceteris paribus, increase the house price growth with 50%.

Table 4.3: Continuous Treatment DID without IPC

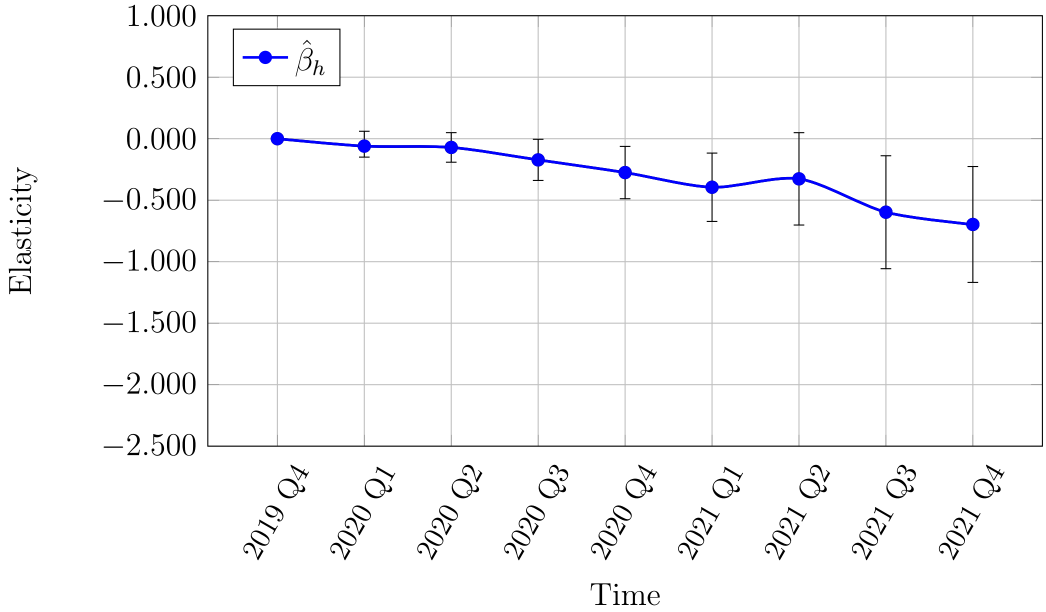
This table provides the coefficient estimate $\hat{\beta}_h$, t-value, F-test, 95% confidence interval - *CI*, R^2 and Adjusted R^2 for the regression done on the quarterly numbers for the amount of headquarters against the house price index in each area. T-statistics are based on robust standard errors. $\hat{\beta}_h$ is a elasticity measure. $\hat{\beta}_h$ represents the difference in house price growth, the greater it deviates from 0, the greater are the differences in house price growth. A negative $\hat{\beta}_h$ indicates that house prices in areas with a larger number of headquarters tend to rise less than house prices in areas with smaller number of headquarters after the fourth quarter of 2019.

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4
$\hat{\beta}_h$	-0.060	-0.071	-0.172	-0.275	-0.395	-0.326	-0.598	-0.697
t-value	-1.33	-1.16	-2.03**	-2.55**	-2.8***	-1.71*	-2.56**	-2.91***
F Statistic	1.76*	1.35	4.1***	6.48***	7.81***	2.92***	6.55***	8.47***
DF	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)
CI	[-0.15:0.029]	[-0.191:0.049]	[-0.339:-0.005]	[-0.488:-0.063]	[-0.673:-0.117]	[-0.702:0.049]	[-1.057:-0.139]	[-1.168:-0.226]
R^2	0.197	0.307	0.445	0.475	0.522	0.562	0.568	0.617
Adj R^2	0.095	0.220	0.375	0.409	0.462	0.506	0.514	0.569

Note: *p<0.1; **p<0.05; ***p<0.01

Figure 4.5: Continuous Treatment DID

This plot shows the coefficient estimate $\hat{\beta}_h$ per quarter from 2019 Q4 to 2021 Q4. The error bars are 95% confidence intervals based on robust standard errors. $\hat{\beta}_h$ represents the difference in house price growth, the greater it deviates from 0, the greater are the differences in house price growth. A negative $\hat{\beta}_h$ indicates that house prices in areas with a high number of headquarters tend to rise less than house prices in areas with a low number of headquarters after the fourth quarter of 2019.



4.1.4 Continuous Treatment DID model with *IPC* as a control variable

We want to control for income based on the same rationale as in the binary treatment regression. When doing so, we get significant results in all regressions. As expected, we are able to increase our coefficient estimates which suggest that the number of publicly traded firms has a more substantial effect on the change in house prices when we control for *IPC*.

Table 4.4: Continuous Treatment DID with IPC

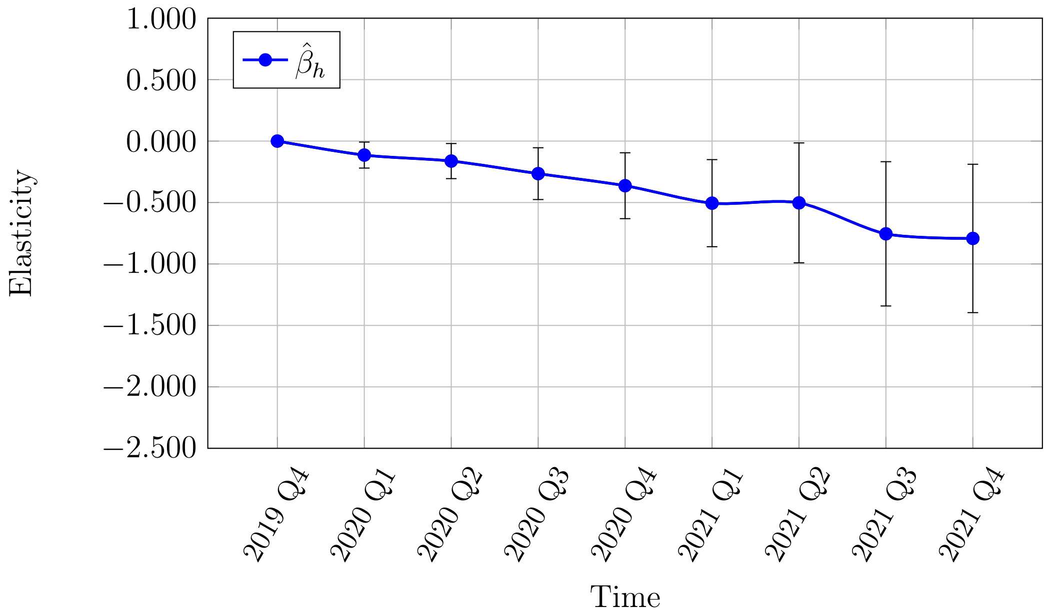
This table provides the coefficient estimate $\hat{\beta}_h$, t-value, F-test, 95% confidence interval - *CI*, R^2 and Adjusted R^2 for the regression done on the quarterly numbers for the amount of headquarters against the house price index in each area. T-statistics are based on robust standard errors. $\hat{\beta}_h$ is a elasticity measure. $\hat{\beta}_h$ represents the difference in house price growth, the greater it deviates from 0, the greater are the differences in house price growth. A negative $\hat{\beta}_h$ indicates that house prices in areas with a larger amount of headquarters tend to rise less than house prices in areas with smaller amounts of headquarters after the fourth quarter of 2019.

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4
$\hat{\beta}_h$	-0.114	-0.163	-0.265	-0.363	-0.505	-0.503	-0.755	-0.792
t-value	-2.12**	-2.24**	-2.47**	-2.66***	-2.8***	-2.03**	-2.53**	-2.58***
F Statistic	2.55**	3.95***	3.07***	3.73***	4.28***	2.06**	3.51***	4.29***
DF	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)
CI	[-0.22:-0.008]	[-0.306:-0.02]	[-0.476:-0.054]	[-0.631:-0.095]	[-0.86:-0.151]	[-0.991:-0.015]	[-1.342:-0.168]	[-1.396:-0.189]
R^2	0.209	0.330	0.456	0.480	0.527	0.568	0.571	0.618
Adj R^2	0.107	0.244	0.385	0.413	0.466	0.512	0.516	0.569

Note: *p<0.1; **p<0.05; ***p<0.01

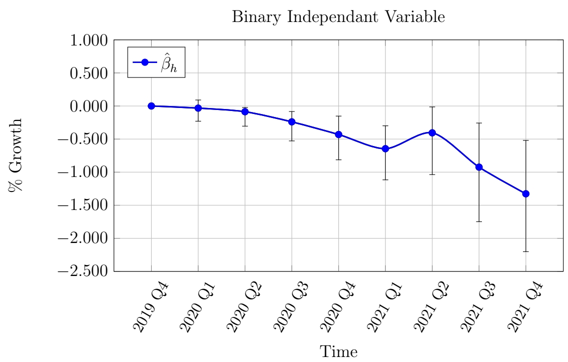
Figure 4.6: Continuous Treatment DID with control for IPC

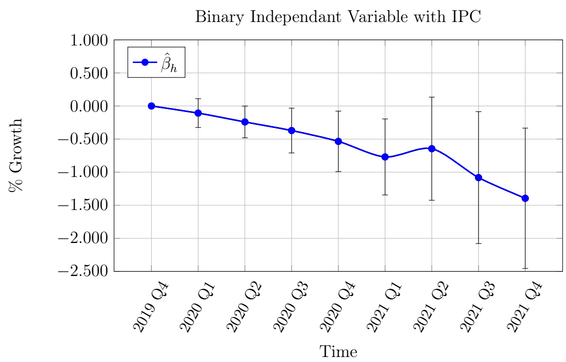
This plot shows the coefficient estimate $\hat{\beta}_h$ per quarter from 2019 Q4 to 2021 Q4. The error bars are 95% confidence intervals based on robust standard errors. $\hat{\beta}_h$ represents the difference in house price growth, the greater it deviates from 0, the greater are the differences in house price growth. A negative $\hat{\beta}_h$ indicates that house prices in areas with a high number of headquarters tend to rise less than house prices in areas with a low number of headquarters after the fourth quarter of 2019.

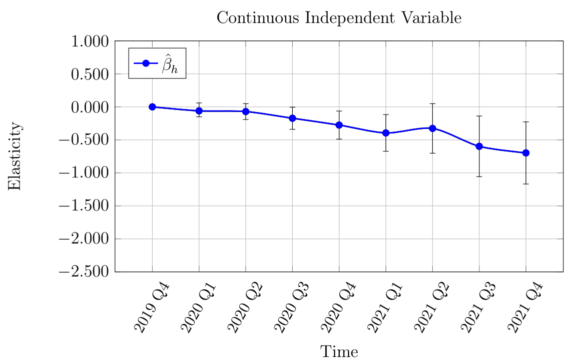


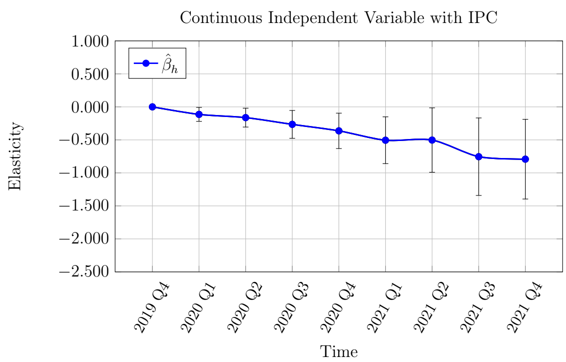
4.1.5 DID Results Summary

Lastly, we summarize our findings presented in the previous sections. Below are the four graphs displayed earlier with its corresponding 95% confidence intervals. One can clearly see how all of our regressions continue to increasingly deviate from zero the further we move from the shock of the fourth quarter of 2019. From the tables we have seen that both R^2 and adjusted R^2 increases as the time after the shock increases. The significance level follows the same pattern, and we observe from the t-test and the f-test that they become more significant as we move away from the fourth quarter of 2019. t-values test the hypothesis that each coefficient is different from 0, while an F-test test whether all the coefficients in the model are different from zero. This suggests that there is a significant difference in house price growth in areas with small number of headquarters versus areas with high number of local headquarters. Before we conclude, it is important to examine the trends and robustness of the results. Following in the next chapter - 4.3 we analyze the results further before our conclusion in chapter 5.









4.2 Robustness

In the following subsection 4.2.1 we present the findings on our pre-trends. Firstly, we compare the below median ZIP codes to the above median ZIP codes. The below median HQ group has a larger HPI growth than the above median group. The reason is logical. Our hypothesis is based on the findings by Choi et al. (2016) who found that areas with a low supply of publicly traded headquarters tend to have their inhabitants investing in investment homes - more so than the inhabitants living in areas with a high concentration of publicly traded firms. The pre-trend growth of the below median areas are therefore in line with the motivation behind our hypothesis. Hence, what needs to be reviewed when comparing the pre-trends to the post shock period is not the pattern itself but if the shock led to a sharper increase in the price growth difference, led by the different savings attitude in below- versus high headquartered areas.

To check if we get a bigger effect in the shock period versus the control period, we compare the coefficient estimates before and after the shock. We start by checking the binary treatment DID regression in Table 4.5. The 7-period lagged regression obtains the strongest coefficient, but decreases before and after that regression interval. On the future logs, we get strong coefficient growth implying a delayed effect. Compared to the pre trends, we get a higher coefficient in regression intervals 7 and 8. We see the same patterns when we control for *IPC*, but get stronger coefficient estimates in the 5-period regression compared to the pre trends, in addition to 7 and 8.

We also compare the pre trends for the continuous treatment DID regression in table 4.5. There seems to be an increase in the effect of headquarters to the change in the HPI, with stronger coefficient estimates in the 4, 5, 7 and 8 period regressions. When we control for income in the continuous treatment DID regression, all regressions but the two period regression yield stronger coefficient estimates than the pre-trends. Further explanation and visualizations are presented in the following section.

4.2.1 Pre Trends

The common practice in difference-in-differences (DID) is to check for parallel trends prior to a treatment. Our treatment group are the areas with a high number of headquarters, and our candidate control group is then the areas with a low number of headquarters. By plotting the *HPI* for both groups we want to observe parallel trends prior to the Covid-19 pandemic. First, we want to illustrate how the house prices have moved for both groups for our whole time period. Figure 1.2 in the introduction illustrated *HPI* for both groups from the first quarter of 1995 until the last quarter of 2021.

Secondly, we want to look at the eight quarters prior to the pandemic against the eight quarters after the pandemic. What we find is in line with what we have described earlier in chapter 4.1.1 i.e., there is some sort of delay effect. We start to observe some difference in the trend 1 to 3 periods after 2019 Q4, but it is not until 2020 that these differences are significant. From Figure 4.1 one can easily see how the actual trend of the areas with a high number of headquarters deviates greatly from the predicted trend. On the other hand, it is not as easy to see what happened before 2019 Q4, cause it seems like there is an ongoing trend here as well. Looking at Figure 4.2 where the fourth quarter of 2019 is set to index 100 it is easier to see how the trend of the increasing growth for the *HPI* on *Low Amount HQ* was somewhat increasing before Covid-19 as well.

In the appendix, we present the pre trend estimates of $\hat{\beta}_h$ for all $h < 0$, and in table 4.5 below we present the corresponding estimated values. We want to take a closer look at whether the differences in house price growth were present before Covid-19, and if so to what extent they were present. This is measured with $\hat{\beta}_h$, and as we can see from the graphs in A2, it may seem that the trend we observed after Covid-19 was also present before. The results of the pre-trends regression are very significant, which in turn supports the hypotheses that there have been a difference in price growth between our two defined areas, this is also what we expected to find based on existing literature.

Table 4.5: Pre-trend DID Regression Summary Table

This table provides the independent variable $\hat{\beta}_h$, t-value, F-test Degrees of Freedom, Confidence interval, R^2 and adjusted R^2 for the four regressions done on the quarterly numbers for the amount of headquarters against the house price index in each area. T-statistics are based on robust standard errors. $\hat{\beta}_h$ and the corresponding confidence interval are presented in percent. $\hat{\beta}_h$ represents the difference in house price growth, the greater it deviates from 0, the greater are the differences in house price growth. $\hat{\beta}_h$ is negative for the periods after our constructed shock period - Q4 2019, and positive for the periods before the shock. A larger number for the pretends indicates that house prices in areas with a larger amount of headquarters tend to rise less than house prices in areas with smaller amounts of headquarters.

	2017 Q4	2018 Q1	2018 Q2	2018 Q3	2018 Q4	2019 Q1	2019 Q2	2019 Q3
Binary regression without IPC								
$\hat{\beta}_h$	0.250	0.601	0.986	0.713	0.577	0.604	0.508	0.303
t-value	0.960	2.45**	4.08***	3.91***	4.08***	5.29***	5.55***	3.66***
F Statistic	0.920	6.02***	16.66***	15.26***	16.66***	28.03***	30.82***	13.38***
DF	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)
CI	[-0.261:0.76]	[0.119:1.083]	[0.299:0.856]	[0.354:1.072]	[0.299:0.856]	[0.379:0.828]	[0.328:0.688]	[0.14:0.466]
R^2	0.641	0.587	0.475	0.500	0.475	0.450	0.334	0.207
Adj R^2	0.595	0.535	0.409	0.437	0.409	0.381	0.251	0.107
Binary regression with IPC								
$\hat{\beta}_h$	0.566	0.759	1.050	0.750	0.538	0.616	0.444	0.243
t-value	2.08**	2.96***	3.36***	4.11***	3.36***	4.73***	4.69***	2.96***
F Statistic	4.33***	4.53***	8.56***	8.64***	8.56***	13.98***	15.95***	7.04***
DF	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)
CI	[0.031:1.102]	[0.255:1.262]	[0.223:0.854]	[0.392:1.108]	[0.223:0.854]	[0.36:0.872]	[0.258:0.63]	[0.081:0.404]
R^2	0.648	0.590	0.475	0.501	0.475	0.450	0.338	0.212
Adj R^2	0.603	0.536	0.407	0.436	0.407	0.379	0.253	0.110
Continuous regression without IPC								
$\hat{\beta}_h$	0.108	0.293	0.448	0.379	0.269	0.233	0.200	0.112
t-value	1.060	2.91***	4.49***	4.3***	4.49***	5.19***	5.23***	3.5***
F Statistic	1.130	8.46***	20.14***	18.49***	20.14***	26.91***	27.35***	12.28***
DF	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)
CI	[-0.092:0.307]	[0.095:0.491]	[0.151:0.388]	[0.205:0.552]	[0.151:0.388]	[0.145:0.322]	[0.125:0.275]	[0.049:0.175]
R^2	0.641	0.589	0.479	0.512	0.479	0.442	0.326	0.200
Adj R^2	0.595	0.538	0.413	0.451	0.413	0.372	0.241	0.099
Continuous regression with IPC								
$\hat{\beta}_h$	0.277	0.395	0.505	0.431	0.265	0.242	0.169	0.080
t-value	2.62***	3.92***	4.09***	4.75***	4.09***	4.54***	4.07***	2.42**
F Statistic	5.52***	8.15***	10.14***	11.29***	10.14***	13.51***	14.17***	6.5***
DF	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)
CI	[0.069:0.484]	[0.197:0.593]	[0.138:0.392]	[0.252:0.609]	[0.138:0.392]	[0.137:0.346]	[0.087:0.25]	[0.015:0.146]
R^2	0.649	0.593	0.479	0.515	0.479	0.442	0.330	0.205
Adj R^2	0.604	0.541	0.412	0.452	0.412	0.370	0.243	0.103

Note: *p<0.1; **p<0.05; ***p<0.01

4.3 Discussion and Limitations

We have based our regression on the House Price Index which is a broad measure of the house prices in the U.S. The data provided by Fannie Mae and Freddie Mac measures the average price changes for houses sold or refinanced. Acting as the dependent variable in our analysis, we measure the index change before and after the shock period. What is important to remember, however, is that our hypothesis is based on whether individuals save in investment homes or take on other investments based on whether they live in areas with a small- or large supply of publicly traded companies. From Choi et al. (2016), we know that primary residents accounted for 67% of the market, vacation properties 12% and investment homes 22%. Since our hypothesis concerns whether the savings shock during the pandemic led to an appreciation of the HPI in the areas with a low supply of publicly traded firms, we are looking at a spark in the investment homes segment of the index. Therefore, the magnitude of our results must be taken into the right context, and considered not only based on the HPI movement itself, but taken into account that the investment homes part of the index is large enough to shift the entire index. Foremost, the continuous treatment provides very large results, which suggest that the number of headquarters have a large effect on the house price growth.

We expect the investment homes market to appreciate in close proximity to the households investing in the unit due to the local bias. Choi et al. (2016) found that households investing in investment homes have a median distance of 25 miles. However, there could be instances where investment homes are bought in different ZIP codes where the households have their primary dwelling. Hence, appreciation in one zip code stemming from investment homes investment can occur due to money inflow from a different ZIP area. Another factor that must be taken into consideration is that investment homes probably have a tendency to be located in cities, with better availability of tenants.

Naturally, one could argue that the savings shock led to an appreciation of not only investment homes but also appreciation of vacation properties and primary residence. We know that the pandemic led to record high vacation homes prices Cororaton (2020), and Choi et al. (2016) also found that in sand states, the investment homes can account for about 11% of the mortgage defaults and excess home price fluctuations. If we assume the primary residence market would fluctuate as normal, and the shock only comes from investment homes and vacation homes, the effect would be substantially higher if we were to isolate the effect of the two latter segments.

Needless to say, the missing opportunity to distinguish the house price appreciation in the different sectors is a clear limitation in our thesis because we want to study a sub-sample of the index, not the entire house price index.

5 Conclusions

In this final chapter, we first provide our conclusions. Second, we emphasise the most important limitations of our analyses and findings. Lastly, we provide a few suggestions for future research.

5.1 Conclusion and Further Research

The local bias toward investing in close proximity to your home is a real life example of imperfect diversification. First, we have shown that the findings of Choi et al. (2016) also apply at the 3-digit ZIP level, and not only for MSAs. There is evidence for an only-game-in-town effect, where investors tend to invest in real estate if there is a lack of publicly traded firms in the area. The phenomenon is closely related to local bias which is described in the article by Choi et al. (2016). Secondly, the savings shock that appeared when Covid-19 hit seems to have appreciated the house price index more than in the pre-trend, especially when we control for income per capita. The effect is increasingly significant, most likely sparked by the hold up period in real estate transactions compared to i.e. a transaction in the stock market.

However, there are omitted variables that can bias the regression results. For instance, controlling for net wealth in addition to IPC could be a good idea when addressing real estate investments. There are a lot more capital constraints on investing in real estate than opening up a brokerage account and buying stocks. The first hypothesis, proven by Van Rooij et al. (2011) suggests that finance literacy is important in predicting who starts saving in the financial markets, but is the lack of public firms in close proximity also a pointer in who has the sufficient financial literacy to invest in the stock market? Another possible explanation that would be interesting to explore is if the reason why investing in local firms is unfavorable is that there are too many investors that drive the price up; i.e. local firms are overpriced.

Naturally, it would also be interesting to go back to our motivation and check whether the phenomenon also applies within Norway and Sweden's multiplicities and counties, i.e. check if a Norwegian county with a low supply of publicly traded firms has a higher appreciation of the house prices than high headquartered counties. The same study can be conducted in Sweden, measuring the implication of our behavioral findings from the introduction.

References

Abdli, M. (2017). På dette området slår både svenskene og danskene oss .

Bennin, R. (1980). Error rates in crsp and compustat: A second look. *The Journal of Finance*, 35(5):1267–1271.

Case, K. E. and Shiller, R. J. (1990). Forecasting prices and excess returns in the housing market. *Real Estate Economics*, 18(3):253–273.

CensusBureau (2019). Per Capita Income. [Online; accessed 2022-05-08].

Choi, H.-S., Hong, H. G., Kubik, J. D., and Thompson, J. P. (2016). Sand states and the us housing crisis. *SSRN 2373179*.

Cororaton, S. (2020). Vacation Home Sales are Getting a Pandemic Boost. [Online; accessed 2022-05-30].

Coval, J. D. and Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance*, 54(6):2045–2073.

Dossche, M. and Zlatanos, S. (2020). COVID-19 and the increase in household savings: precautionary or forced? *Economic Bulletin Boxes*, 6.

FHFA (2022). FHFA House Price Index Federal Housing Finance Agency. [Online; accessed 2022-05-29].

French, K. R. and Poterba, J. M. (1991). Investor diversification and international equity markets. *The American Economic Review*, 81(2):222–226.

Iachan, F. S., Nenov, P. T., and Simsek, A. (2021). The choice channel of financial innovation. *American Economic Journal: Macroeconomics*, 13(2):333–72.

IMF (2021). Imf Global Housing Watch. [Online; accessed 2022-05-28].

Lewis, K. K. (1999). Trying to explain home bias in equities and consumption. *Journal of economic literature*, 37(2):571–608.

Mian, A. and Sufi, A. (2009). The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis*. *The Quarterly Journal of Economics*, 124(4):1449–1496.

Polkovnichenko, V. (2004). Limited stock market participation and the equity premium. *Finance Research Letters*, 1(1):24–34.

Rambachan, A. and Roth, J. (2019). An honest approach to parallel trends. *Unpublished manuscript, Harvard University.[99].*

Rieger, J. (2017). Limited market participation, housing and asset prices.

Smarty (2022). ZIP Code Zones Map. [Online; accessed 2022-05-07].

SP (2022). Sp corelogic case-shiller home price indices methodology. *SP Dow Jones Indices*, 1(1):24–25.

Spielman, S. E., Folch, D., and Nagle, N. (2014). Patterns and causes of uncertainty in the american community survey. *Applied Geography*, 46:147–157.

SSB (2022). Ssb Prisindeks Brukte Boliger. [Online; accessed 2022-05-28].

Van Rooij, M., Lusardi, A., and Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial economics*, 101(2):449–472.

Appendices

A1 DID Summary Estimates

Table A1.1: Regression Summary Table

This table provides the independent variable $\hat{\beta}_h$, t-value, F-test, 95% confidence interval - *CI*, R^2 and Adjusted R^2 for the four regressions done on the quarterly numbers for the amount of headquarters against the house price index in each area. T-statistics are based on robust standard errors. $\hat{\beta}_h$ is presented in percent. $\hat{\beta}_h$ represents the difference in house price growth, the greater it deviates from 0, the greater are the differences in house price growth. A negative $\hat{\beta}_h$ indicates that house prices in areas with a larger amount of headquarters tend to rise less than house prices in areas with smaller amounts of headquarters after the fourth quarter of 2019.

	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4
Binary regression without IPC								
$\hat{\beta}_h$	-0.032	-0.087	-0.240	-0.431	-0.645	-0.406	-0.924	-1.327
t-value	-0.32	-0.78	-1.64	-2.22**	-2.7***	-1.26	-2.2**	-2.98***
F Statistic	0.1	0.61	2.68***	4.92***	7.26***	1.6	4.85***	8.9***
DF	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)
CI	[-0.23:0.166]	[-0.305:0.132]	[-0.527:0.048]	[-0.813:-0.049]	[-1.116:-0.174]	[-1.037:0.225]	[-1.749:-0.099]	[-2.202:-0.452]
R^2	0.192	0.304	0.438	0.466	0.513	0.557	0.561	0.612
Adj R^2	0.090	0.216	0.367	0.399	0.451	0.501	0.505	0.563
Binary regression With IPC								
$\hat{\beta}_h$	-0.107	-0.241	-0.371	-0.535	-0.770	-0.645	-1.082	-1.395
t-value	-0.97	-1.98**	-2.16**	-2.3**	-2.64***	-1.63	-2.14**	-2.59***
F Statistic	1.13	3.79***	2.38**	2.79***	3.89***	1.33	2.55**	4.45***
DF	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)
CI	[-0.325:0.11]	[-0.481:-0.001]	[-0.71:-0.032]	[-0.992:-0.077]	[-1.344:-0.195]	[-1.424:0.134]	[-2.079:-0.085]	[-2.456:-0.334]
R^2	0.198	0.321	0.443	0.468	0.514	0.560	0.561	0.612
Adj R^2	0.095	0.233	0.372	0.400	0.452	0.503	0.505	0.562
Continuous regression without IPC								
$\hat{\beta}_h$	-0.060	-0.071	-0.172	-0.275	-0.395	-0.326	-0.598	-0.697
t-value	-1.33	-1.16	-2.03**	-2.55**	-2.8***	-1.71*	-2.56**	-2.91***
F Statistic	1.76*	1.35	4.1***	6.48***	7.81***	2.92***	6.55***	8.47***
DF	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)	(1, 357)
CI	[-0.15:0.029]	[-0.191:0.049]	[-0.339:-0.005]	[-0.488:-0.063]	[-0.673:-0.117]	[-0.702:0.049]	[-1.057:-0.139]	[-1.168:-0.226]
R^2	0.197	0.307	0.445	0.475	0.522	0.562	0.568	0.617
Adj R^2	0.095	0.220	0.375	0.409	0.462	0.506	0.514	0.569
Continuous regression with IPC								
$\hat{\beta}_h$	-0.114	-0.163	-0.265	-0.363	-0.505	-0.503	-0.755	-0.792
t-value	-2.12**	-2.24**	-2.47**	-2.66***	-2.8***	-2.03**	-2.53**	-2.58***
F Statistic	2.55**	3.95***	3.07***	3.73***	4.28***	2.06**	3.51***	4.29***
DF	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)	(2, 356)
CI	[-0.22:-0.008]	[-0.306:-0.02]	[-0.476:-0.054]	[-0.631:-0.095]	[-0.86:-0.151]	[-0.991:-0.015]	[-1.342:-0.168]	[-1.396:-0.189]
R^2	0.209	0.330	0.456	0.480	0.527	0.568	0.571	0.618
Adj R^2	0.107	0.244	0.385	0.413	0.466	0.512	0.516	0.569

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

A2 Pre-trend Illustrations

Figure A2.1: Binary Independent Variable Over Time without IPC

This plot shows the independent variable $\hat{\beta}_h$ per quarter from 2017 Q4 to 2021 Q4. The more $\hat{\beta}_h$ deviates from 0, the higher are the differences between house price changes in areas with above median amount of headquarters ($Zip_{i,Above}$) against areas with below median amount of headquarters ($Zip_{i,Below}$). All values are presented in percent. 2019 Q4 is artificial plotted as 0 since this is the period of our proposed shock. $\hat{\beta}_h$ is negative for the periods after our constructed shock period - Q4 2019, and positive for the periods before the shock.

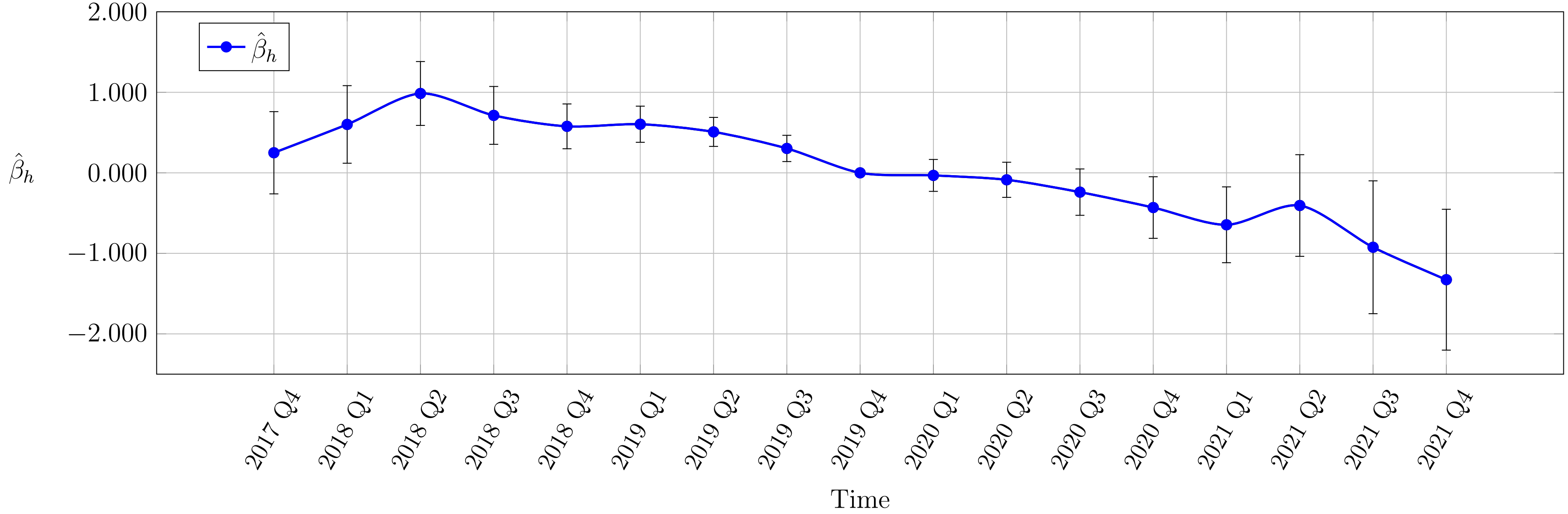


Figure A2.2: Binary Independent Variable Over Time with IPC

This plot shows the independent variable $\hat{\beta}_h$ per quarter from 2018 Q1 to 2021 Q4. The error bars are 95% confidence intervals based on robust standard errors. The more $\hat{\beta}_h$ deviates from 0, the higher are the differences between house price changes in areas with above median amount of headquarters ($Zip_{i,Above}$) against areas with below median amount of headquarters ($Zip_{i,Below}$). All values are presented in percent. 2019 Q4 is artificial plotted as 0 since this is the period of our proposed shock. $\hat{\beta}_h$ is negative for the periods after our constructed shock period - Q4 2019, and positive for the periods before the shock.

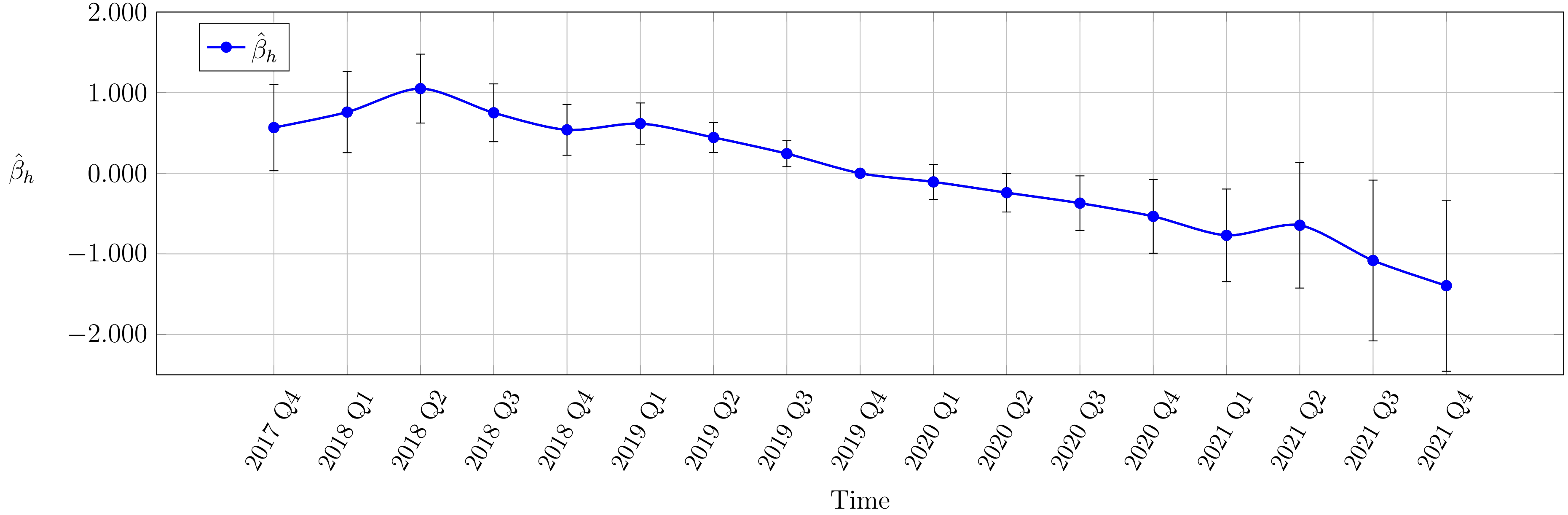


Figure A2.3: Continuous Independent Variable Over Time with IPC

This plot shows the independent variable $\hat{\beta}_h$ per quarter from 2017 Q4 to 2021 Q4. The error bars are 95% confidence intervals based on robust standard errors. The more $\hat{\beta}_h$ deviates from 0, the higher the difference between house price changes in areas with a large amounts of headquarters against areas with low amount of headquarters. All values are presented in percent. 2019 Q4 is artificial plotted as 0 since this is the period of our proposed shock. $\hat{\beta}_h$ is negative for the periods after our constructed shock period - Q4 2019, and positive for the periods before the shock.

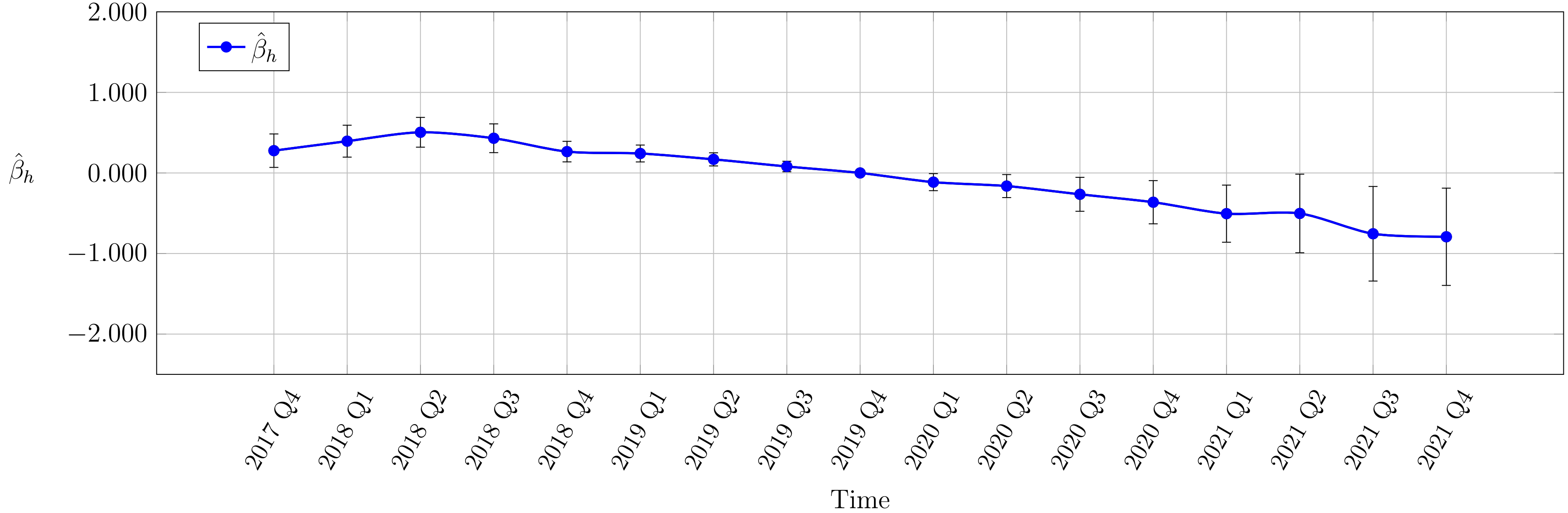
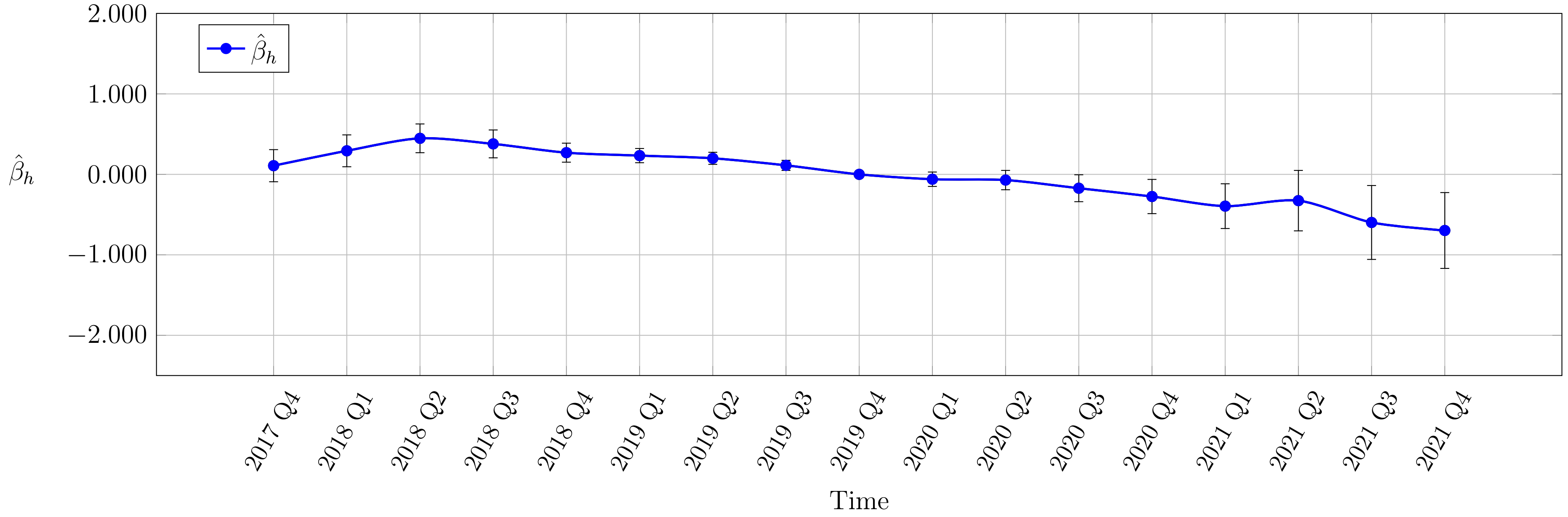
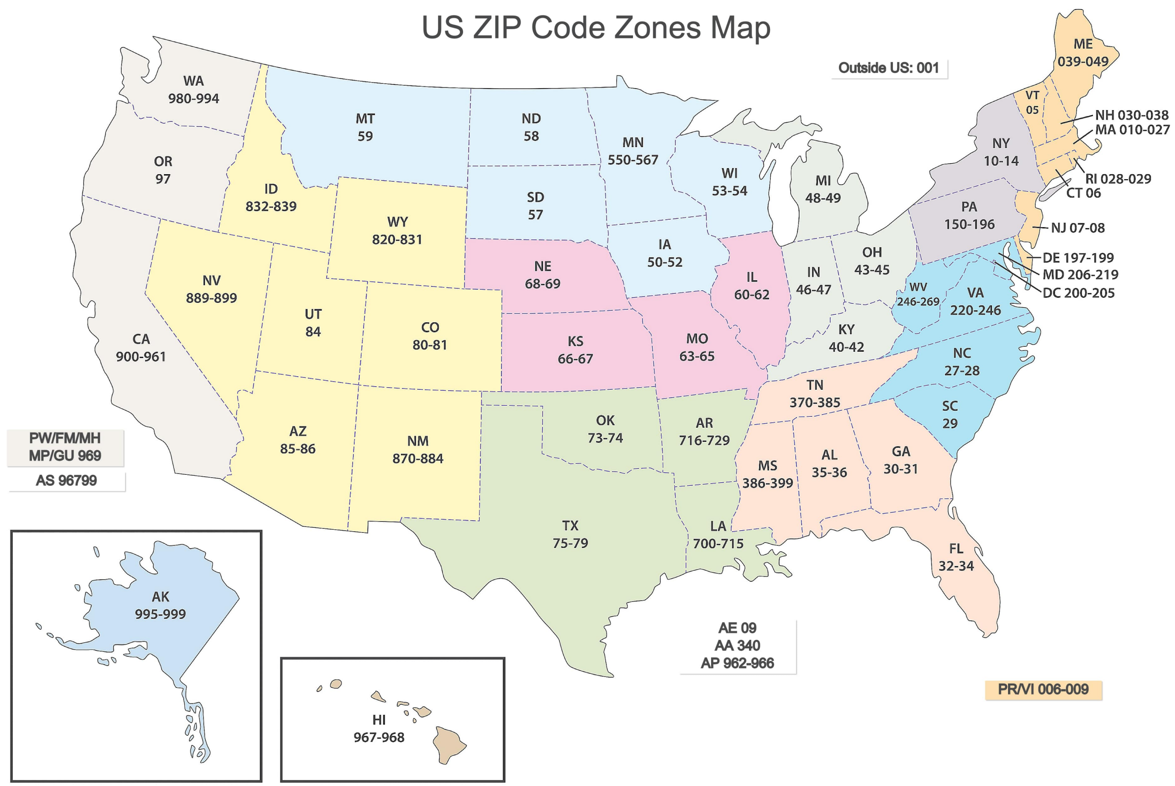


Figure A2.4: Continuous Independent Variable Over Time without IPC

This plot shows the independent variable $\hat{\beta}_h$ per quarter from 2017 Q4 to 2021 Q4. The error bars are 95% confidence intervals based on robust standard errors. The more $\hat{\beta}_h$ deviates from 0, the higher the difference between house price changes in areas with a large amount of headquarters against areas with low amounts of headquarters. All values are presented in percent. 2019 Q4 is artificial plotted as 0 since this is the period of our proposed shock. $\hat{\beta}_h$ is negative for the periods after our constructed shock period - Q4 2019, and positive for the periods before the shock.



A3 US ZIP Code Zones Map



Source: (Smarty, 2022)