Regional Variation in U.S. Homeownership Rates

Shikha Agarwal

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Eric Zivot, Chair

Thomas Gilbert, Chair

Philip Brock

Program Authorized to Offer Degree:

Economics

University of Washington

Abstract

Regional Variation in U.S. Homeownership Rates

Shikha Agarwal

Chair of the Supervisory Committee:

Eric Zivot

Department of Economics

Thomas Gilbert

Department of Finance and Business Economics

Homeownership rate varies dramatically across states, metro areas and rural and urban areas within the United States. While cross-country homeownership differences can be explained by national differences in history, institutions, and macroeconomic conditions, these factors cannot explain why homeownership rates vary across regions, given that regions in the same country share similar historical, institutional, and macroeconomic arrangements. This study aims to explain the cross-sectional and inter-temporal differences in homeownership between Census

regions using zip code level data covering approximately 90 percent of all zip codes in the country. Specifically, do the factors that explain intra-regional variation in homeownership rate have similar effects across regions and, what is the differential impact of housing market variables related to home value, rent and household income across different regions? I conclude that while relationship between demographic factors and homeownership is consistent for all zip codes regardless of region, housing market variables like price-rent ratio and price-income ratio have different relationship with homeownership across zip codes depending on the regional location. This study also investigates the impact of the housing bubble on homeownership during the bust and the initial phase of economic recovery on three dimensions. Is there a structural break in the national housing market in 2011-2012? Did regional disparities due to unobserved region-specific factors significantly decrease during the recovery period of 2012-2016? Finally, since most of the run-up in homeownership was due to relaxed credit standards, how did the response of homeownership to relative affordability change before and after the crash in house prices? The Chow test concludes that housing markets were indeed quite different before and after the housing crash at both the national and sub-national level. Not only did the homeownership response change for all included variables and for all regions, but also the recovery phase can better explain collective variation across regions and across zip codes. Results also show that the homeownership response of households to relative affordability became stronger due to both personal experience of foreclosures as well as tighter lending standards.

Introduction

Shikha Agarwal

The rate of homeownership is traditionally interpreted as a major indicator of economic welfare. Housing services consumption alone was 12% of GDP as of the third quarter of 2016, nearly 100 basis points below the 30-year average[1]. Residential fixed investment is another category of housing related expenditure which contributes to GDP. The promotion of homeownership has for decades been one of the most popular political objectives in the United States.

The decade through 2004 witnessed robust economic expansion and a policy environment promoting broad access to homeownership. For instance, the Clinton Administration and the Bush Administration enacted policies to have lower downpayment requirements, and Zero-Downpayment Initiative for first time homebuyers, respectively[2]. The gains in homeownership started eroding with bursting of the housing bubble of 2007. Subsequently, in the aftermath of the ensuing recession, homeownership fell with tighter credit conditions, rising student loan debt, stagnant real incomes, and perhaps a subtle change in attitudes toward homeownership.

A significant part of the existing empirical research focuses either on microeconomic decisions of owning versus renting, or on cross-country differences in homeownership rates. While cross-country homeownership differences can be explained by national differences in history, institutions, and macroeconomic conditions, these factors cannot explain why homeownership rates vary across regions, given that regions in the same country share very similar historical, institutional and macroeconomic arrangements.

[1] Rosen et al. (2017)
[2] Chambers et al. (2009)

Homeownership rates in different regions within the United States have exhibited large independent movements despite facing symmetric interest rates and federal tax treatment[3]. The West experienced a steady decline in homeownership rates from 1967 through the early 1990s. As the Northeast's rate marched upwards, the Midwest and South regions experienced extended stagnation in homeownership rates. Regional disparities within the United States underscore the role of local factors in influencing the homeownership rate.

Regional housing prices also react uniformly to certain national economic factors, such as mortgage rates. On the other hand, local factors such as population shifts, employment and income trends often have unique impact on housing prices. During the housing crisis of 2007-2008, all regions witnessed different extent of decline in homeownership rate. The impact on home prices and the extent to which a bubble developed also depended largely on local conditions. While the Western region had the largest concentration of foreclosures and the sharpest declines in home prices, the Northeast witnessed the smallest decline in housing prices and homeownership rate amongst the four regions.

These experiences indicate that variation in homeownership rates must be analyzed by superimposing broad national trends upon unique regional trends instead of assuming a single homogeneous national housing market. There is a huge variation across the nation, with some states, particularly in the middle of the country, having much higher rates than others. There is also a difference between metropolitan and nonmetropolitan areas, with non-metro areas generally having higher homeownership rates. Finally, certain expensive cities on the coasts have homeownership rates that are lower than both their state and other metro areas. Explaining this variation is a promising topic for future study [4].

Moreover, the issue of regional disparity has largely been discussed using data from the cities or the metro areas, while the rural areas have been left out of such studies. Such datasets are not only smaller but oversample highly regulated cities and underrepresent the bulk of American housing. Hence, analyzing the regional variabil-

[3] Painter and Redfearn (2002)
[4] Goodman and Mayer (2018)

ity of homeownership rates using a more inclusive dataset than just the metro areas is vital to gaining more insight into the regional economic and socio-demographic factors affecting household tenure choice.

Multiyear estimates from the American Community Survey (ACS) produces data on 35 housing variables amongst other key characteristics at extremely low geographic levels. Such estimates offer increased statistical reliability of the data for less populated areas and small population subgroups. I use the aggregate 5-year estimates from the ACS at the zip code level covering approximately 90% of all zip codes in the country. I assign region membership to each zip code using the state Federal Information Processing Standards (FIPS) codes.

The first chapter focuses on the cross sectional variation in homeownership rates across zip codes and across regions during the 2007-2011 collapse in housing prices, using the 5-year ACS estimates for this period. Specifically, do the factors that explain intra-regional variation in homeownership rate have similar effects across regions? I also investigate the differential impact of housing market variables related to home value, rent and household income across different regions. I use six explanatory variables, three dummy variables and relevant interaction terms in a linear regression model to explain the variation in homeownership rates. Households' social or demographic characteristics may reflect their preference towards owning or renting.

According to the 2007 Survey of Consumer Finances (SCF), the median homeowner is more educated, has higher earnings and higher wealth, and is older than renters. Older households are more likely to be homeowners, because age is a proxy for wealth, and wealthy households are more likely to own homes. Similarly married couples are much more likely to be homeowners than either those living alone or single households living with other relatives. Education is usually seen to be representative of the expected lifetime earnings of an individual. Increased earnings potential from additional education makes saving for a down payment, qualifying for a mortgage, and affording monthly mortgage payments easier. A higher level of education is correlated with higher credit scores, higher financial literacy and better comprehension of loan options.

Housing units in multifamily structures are typically characterized by higher den-

sity and urban living. These are also more often associated with rental occupants than owner occupants. Lack of maintenance, ease of relocation, and convenience in terms of location and transportation may encourage households to occupy rental multifamily housing. Based on economic theory, demographic factors like age and education level of the household head, household structure and multifamily housing in a zip code should have similar effects across regions. However, housing market variables like price-rent ratio and price-income ratio have different impact in different regions. Price-rent and price-income ratios that would be considered "high" for one region may be typical for another.

I find that demographic variables and housing density in an area are important in explaining homeownership rate in a zip code and bear the expected relationship with homeownership. However, the four Census regions should be seen as four different housing markets since homeownership is related to key housing market variables in different ways across regions. I find that at the national level, price-rent ratio is negatively related to the homeownership rate. Since zip codes within a region are expected to be similar, I find a positive relationship between price-rent ratio and homeownership for the Midwest and the Northeast, and near-zero for the West and the South. Similarly, the effect of price-income ratio on homeownership is very different in magnitude for zip codes located in different regions.

While the first chapter focuses only on one cross-sectional analysis, the second chapter uses a panel data for two time periods to analyse changes in the housing market and homeownership behavior during and after the Great Recession. I use the aggregate 5-year data from 2011 and 2016 ACS at the zip code level. My choice of time periods is based on the developments of the housing bubble and the resulting credit crisis. Home prices reached their peak in the second quarter of 2006 and did not start recovering until the end of 2011. The first time period of 2007-2011 is the bust phase in the housing market and the second period of 2012-2016 can be thought of as the phase of initial recovery. I utilize the best model developed in the first chapter to proceed with my analysis.

Assuming that the regional housing markets changed after the crisis of 2007-08, I evaluate the econometric model for both time periods and test for joint equality

of the coefficients using the Chow test. The Chow test rejects the null hypothesis of no structural break or parameters shift at a high level of statistical significance. As expected, I find that the relationship of homeownership to the variables included in my model is indeed significantly different in the recovery period relative to the crisis period. The housing crisis not only altered lending and banking practices but also households' behavior and attitude towards homeownership.

The idea of a return to more 'normal' times is also analysed by testing for regional differences in homeownership due to unobserved factors. More specifically, I create a measure of collective variation and the reduction in this measure across time to build a test statistic. I utilize the basic nonparametric bootstrap algorithm to generate distribution of this custom test statistic and test for a significant difference of means using a paired t-test. I find that included factors have a more homogeneous effect on the homeownership rate across regions in the recovery period than in the crisis period. In other words, the collective variation across regions and across zip codes can be explained more by fundamentals in the recovery than in the bust.

Next, I focus on the response of homeownership rate in a zip code to the ratio of median home value to median household income. The homeownership rate increased during the housing boom through the mid-2000s, fueled by an extremely loose credit environment and sustained home price appreciation. U.S. housing and mortgage markets became stressed during 2007 and 2008 as a result of significant house price declines and the weakening economy. Following the end of the Great Recession, new rules and standards on lenders contributed to the constrained credit environment and substantially inhibited lending even to the most creditworthy households. Consistent with my hypothesis, I find that homeownership is more "tightly" associated with relative affordability in the recovery period. This can be attributed to a dramatic shift from lax lending standards to a much tighter credit regime.

This research improves upon the existing empirical studies in three main ways. Firstly, existing studies are based on data from individual households or on aggregate data at the county or the statistical area level. No investigation of the ownership outcomes across regions and across time has used aggregate data for a large number of zip codes. In contrast, my study uses aggregate data at the zip code level from

approximately 90% of all zip codes covered by the ACS. This granularity of data also accounts for heterogeneity across submarkets within the county or the statistical area.

Secondly, the issue of regional disparity has largely been discussed using data from the cities or the metro areas and the rural areas have largely been left out. Such datasets may oversample highly regulated cities and underrepresent the bulk of American housing. My study uses the most comprehensive dataset with almost 40% zip codes belonging to counties where more than half the population lives in rural areas. Approximately 64% of zip codes in my dataset do not belong to a Metropolitan Statistical Area (MSA) and at least 50% zip codes are completely rural.

Lastly, I create a custom metric and test statistic to confirm the change in the housing market at the regional level. This serves as a way to isolate regional differences from zip code level data and focus on only collective inter-regional disparity in homeownership rates that exists due to unobserved or excluded factors. This research is a novel attempt to quantify the difference in homeownership rates, conditional on the included variables. In the absence of any regional variation apart from what can be accounted for demographic and market variables, this measure should be statistically close to zero. This metric is then compared across the bust and recovery periods to yield insights into the changes that occurred in the housing market due to the Great Recession.

While the specific empirical results pertain only to the included zip code level sample and time period of 2007-2016, the presence of this regional variation merits a more general discussion. Both the coastal regions, West and Northeast, have the lowest homeownership rates, highest median household incomes and highest median home values. Much of this can be attributed to the various forms of supply side constraints in these regions. High-slope and coastal areas are severely land-constrained by their topography, rendering housing supply relatively inelastic. Furthermore, the extent of topographical constraints correlates positively and strongly with regulatory barriers to development.

Land use regulations in the United States are widespread, largely under local control, and may be a major factor accounting for why housing appears to be in inelastic supply in many of the larger coastal markets. Zoning ordinances enacted

by local governments greatly affect the costs, size and quantity of new single family homes built. In areas with stricter regulations, building costs typically increase and fewer homes are built, especially when compared with areas with fewer regulations. As land use regulations stifle new supply, land prices and construction costs rise, greatly reducing affordability and limiting homeownership. Housing supply factors are relatively harder to quantify and incorporating these is beyond the scope of this study. These supply side constraints are, however, likely to manifest in the observed differential relationship between homeownership and housing prices variables.

The existing level differences in regional dummies may be due to other unobserved factors that are likely to generate heterogeneous inter-regional association with homeownership. For instance, average mortgage rates, household's marginal tax bracket and occupation profile may vary depending on the location. Spatial patterns of subprime lending reveal a disproportionately large market share in low income, low education and minority neighborhoods. Similarly, declines in household income observed during the Great Recession correlate to the type of jobs lost. Composition of housing stock, cost of maintenance and depreciation and cultural preferences towards homeownership are some of the other unobserved factors that may either confound with observed determinants or have independent effects.

Summarizing, my research improves our understanding of variation in homeownership rates across regions and over time. The four regions should be considered as four distinct housing markets, and this has important policy implications. Additionally, understanding the interaction between different regional housing markets may improve forecasts of home values. My dissertation, though leaves many questions unanswered, outlines a road map using publicly available data, provides baseline econometric framework and test statistic for future work and identifies opportunities for improvement.

Chapter 1: Cross-sectional Heterogeneity in U.S. Homeownership Rates

Shikha Agarwal

Abstract

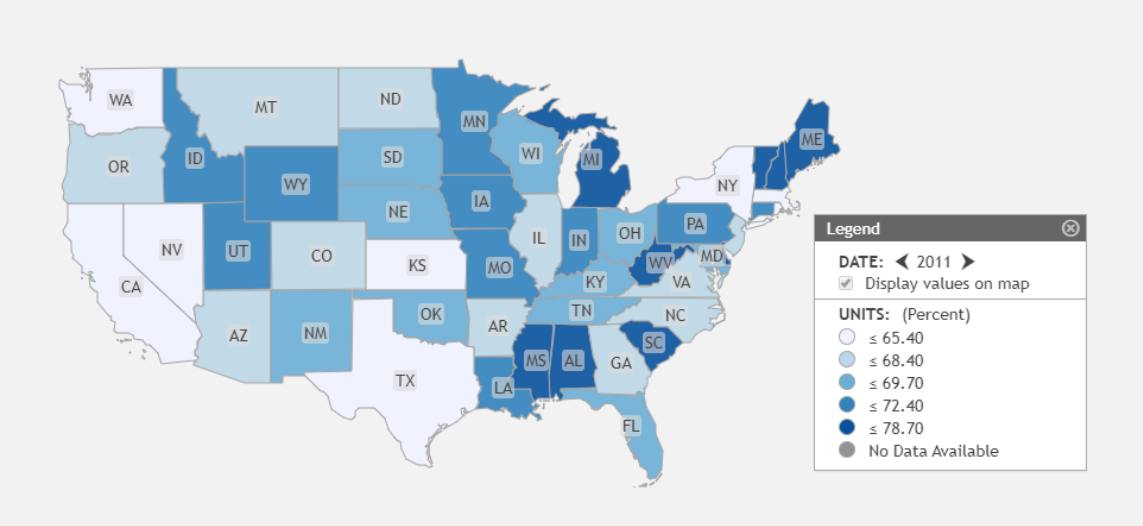
The issue of regional disparity in homeownership rates has largely been discussed using data from the cities or the metro areas and the rural areas have been left out of such studies. I use U.S. zip code level data to explain differences in homeownership rates across Census regions. Specifically, do the factors that explain intra-regional variation in homeownership rate have similar effects across regions? I find that age and education of household have the same positive relationship with homeownership across all regions. Similarly, homeownership is seen to be lower for non-family households and in densely populated urban areas. I find that this relationship is also consistent across regions. I also investigate, what is the differential impact of housing market variables related to home value, rent and household income across different regions? I find that the four regions can be considered as four separate housing markets. The regression coefficient on price-rent ratio for the Midwest is more than three times that for Northeast. Similarly, the ratio of median home value to median household income price-income ratio which is an indicator of relative affordability in a zip code has very different impact on homeownership depending on the regional location. South is the only region where price-income ratio is associated positively with homeownership rate, which is inconsistent with my hypothesis.

1 Introduction

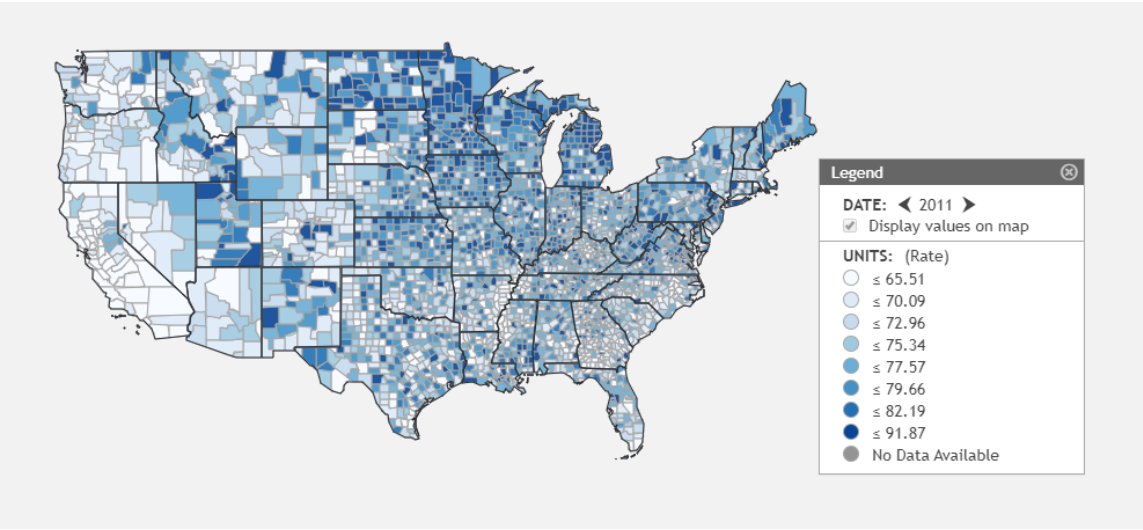
The rate of homeownership is traditionally interpreted as a major indicator of economic welfare. Together, housing services and residential fixed investment spending constitute the majority of housing related expenditure nationwide, and account for more than 15% of GDP. Beyond improving overall economic conditions, homeownership is beneficial to both individual households and society as a whole. Homeownership is associated with a high level of wealth accumulation and economic freedom (Turner and Luea (2009)). Homeowners also tend to be more satisfied with their housing situation than tenants (Elsinga and Hoekstra (2005)). Several studies even report a positive influence of homeownership on social cohesion and stability (Rossi and Weber (1996), DiPasquale and Glaeser (1999), Dietz and Haurin (2003)). It has been reported that, relative to renting, homeownership generates neighborhood benefits related to property upkeep, public safety, school quality, and the like (Green et al. (1997), Aaronson (2000)).

A significant part of the existing empirical research focuses either on microeconomic decisions of owning versus renting, or on cross-country differences in homeownership rates (Proxenos (2002), Fisher and Jaffe (2003), Earley (2004), Gwin and Ong (2008)). Homeownership rates differ vastly across Europe, ranging from over 80% in Spain and Ireland to around 40% in Germany and below 35% in Switzerland. Within Germany itself, regional homeownership rates vary as much as 50 percentage points. Within Switzerland, there exist large difference between the cantons.

There is, however, a surprising paucity of studies empirically analyzing differences in owner-occupation rates within countries. While cross-country homeownership differences can be explained by national differences in history, institutions, and macroeconomic conditions, these factors cannot explain why homeownership rates vary across regions, given that regions in the same country share very similar historical, institutional and macroeconomic arrangements. Analyzing the regional variability of homeownership rates hence creates an opportunity to gain more insight into the regional economic and socio-demographic factors affecting household tenure choice.



(a) Variation in Homeownership Rate across States



(b) Variation in Homeownership Rate across Counties

Figure 1: **Cross-sectional Variation in U.S. Homeownership Rate**

Source: https://geofred.stlouisfed.org/map
Note 1: Panel (a) displays annual average homeownership rate for states in contiguous U.S. in 2011. The states of Alaska and Hawaii fall in the lowest homeownership rate category. States are divided into five fractiles of homeownership rate, with each class consisting of 9-11 states. Values range from 44.8% to 78.7%.
Note 2: Panel (b) displays 5-year estimate of homeownership rate for counties in contiguous U.S. during the period 2007-2011. Counties are divided into eight fractiles of homeownership rate, with each class consisting of 392-394 counties. Values range from 0.0% to 91.9%.

Figures 1a and 1b show the extent of variation in homeownership rates across states and across counties, respectively. A deeper color represents a higher homeownership rate. At the state level, the average homeownership rate in 2011 ranges from less than 65.4% to 78.7%. This variation gets even more striking at the county level. Based on a 5-year estimate of homeownership rates during 2007-11, the range is from less than 65.5% to 91.8%. Similarly, there is also a difference between metropolitan and micropolitan areas, with non-metro areas generally having higher homeownership rates. Figures 1a and 1b indicate that we expect even more variation at a more sub-aggregated level of geography like zip codes. A closer look at figure 1b also brings out that most counties in the western part of the country fall in the bottom two fractiles of distribution of homeownership rate. In general, areas within a region are expected to be more similar than areas in another region (Lee and Myers (2003)).

I aim to investigate the factors that explain cross sectional variation in homeownership rate across zip codes and across regions. Specifically, do the factors that explain intra-regional variation in homeownership rate have similar effects across regions? By focusing on households that own both their principal residence and additional real estate, Ioannides and Rosenthal (1994) confirm that the investment and consumption demands for housing differ. Specifically, the investment demand for housing is more sensitive to wealth and income than is the consumption demand for housing, but consumption demand is more sensitive to demographic variables like age, education, and family size.

I also investigate, what is the differential impact of housing market variables related to home value, rent and household income across different regions? Painter and Redfearn (2002) emphasize that regional disparities within the United States underscore the role of local factors in influencing homeownership rate. The West experienced a steady decline in ownership rates from 1967 through the early 1990s as New England's rate marched upwards. The Midwest and South regions experienced extended stagnation in ownership rates, punctuated by infrequent but significant changes. The note that all of this intra-regional variation occurred under generally uniform interest rate and federal tax regimes.

I use the aggregate 5-year data from 2011 American Community Survey (ACS)

at the zip code level. These contain 'period' estimates of data collected over the time period 2007-2011. My sample contains 29,851 zip codes consisting of both urban and rural areas. Additionally, I assign the regional location to each zip code based on U.S. Census codes. I use six explanatory variables, three dummy variables and relevant interaction terms to explain variation in homeownership rates across regions. The region can affect the homeownership rate of an area not only through its unobserved locational effect but also through interaction with market variables. Three dummy variables are used in a linear regression model with intercept to avoid dummy variable trap with four regions. I choose 'West' as the omitted category since it has the lowest homeownership rate throughout my time period of interest, 2007-2011.

The results are mostly consistent with my hypotheses. I find that demographic and density variables have the expected signs and high statistical significance. Based on economic theory of housing tenure choice, there is no reason to expect differential effect of these variables. However, I also find that each region is characterised by different housing market conditions. Price-to-rent and price-to-income are two commonly considered housing market indicators. My results show that homeownership responds very differently to these indicators across regions. The regression coefficient of homeownership rate on price-rent ratio is positive and largest in magnitude for the Midwest. I conclude by showing that my included variables indeed reduce the differences between regions that can be owed to unobserved factors.

This paper improves upon the existing empirical studies in two main ways. Firstly, existing studies are based on a data from individual households or on aggregate data for only Metropolitan Statistical Areas (MSA). No investigation of the ownership outcomes has used aggregate data for a large number of zip codes. My study takes into account the heterogeneity across submarkets within MSAs, not just across MSAs. Eilbott and Binkowski (1985) use aggregate MSA data from the 1970 Census to explain homeownership rates and find that over 56% of the variation can be explained by the age distribution of the households. Coulson (2002) uses a standard probit model of the individual homeownership decision where the micro-level observations are aggregated at the regional level to determine the causes of inter-regional

disparities. Secondly, the issue of regional disparity has largely been discussed using data from the cities or the metro areas and the rural areas have been left out of such studies. In fact, Glaeser and Gyourko (2002) point out that smaller datasets which feature only large metropolitan areas may oversample highly regulated cities and underrepresent the bulk of American housing.

The remainder of the paper is organized as follows. Section 2 reviews general literature and section 3 develops the specific hypotheses of interest. Section 4 describes the data source and sample construction. Section 5 describes the econometric specifications and model selection. Section 6 presents the results and section 7 provides concluding remarks.

2 Literature Review

The demand for housing services is determined by socioeconomic characteristics, such as income (determined in part by skills), age, and household size, with which households are "endowed." After the level of demand for housing services is determined, the user cost of owning relative to renting (Goodman (1988); Henderson and Ioannides (1983)) provides a framework to analyze tenure choice. According to Zhou (2013), expectations influence people's tenure decisions. If a person expects the value of a house, the cost of renting, or his income to increase, he will be more likely to buy a house, and vice versa. However, the impact of expectation is short-term because people will change their expectations based on their information set about the housing market. For an example, a young household that prefers to buy a house in their 30s may delay entering owner-occupation for a few years because they expect price to fall.

Preference, on the contrary, may change people's decision in the long-run. A young household that prefers to rent in their 30s may not buy a house until their 40s. People's social or demographic characteristics may reflect their preference towards owning or renting. These factors, include age, education, occupation, marital status, race and immigration status, could change the way that people think about the tenure decision. Thus, if young households in different time have different char-

acteristics, or value these characteristics differently, their preference for tenure choice may also change in the long-term. Many studies have focused on the factors that reveal people's expectations. These factors include the owning cost (such as down payment, mortgage payment, interest rate, property tax, and maintenance cost) or renting cost (mainly rent), housing (either owning or renting) cost relatively to income, and potential capital gain from owning. By comparing the current level of these factors to the previous level, people will form expectations about future, and make tenure choice accordingly.

According to the 2007 Survey of Consumer Finances, the median renter is 39 years old, while the median homeowner is older, 52. 40% of homeowners have a college degree but only 25% of renters. Earnings for the median homeowner are about twice as high as those for the median renter, $51,418 versus $24,887. Homeowners are also wealthier: the median wealth-to-earnings ratio for homeowners is almost 4 but a meager 0.21 for renters. Finally, homeowners hold a staggering 97.5% of aggregate household wealth in the United States. Briefly, the median homeowner tends to be more educated, has higher earnings and higher wealth, although on average homeowners are older than renters.

The empirical literature also finds that wealth constraints are typically more important than income constraints when purchasing a home; see, for instance Linneman et al. (1997) or Quercia et al. (2003). In a world with credit market imperfections, one of the main benefits from homeownership is that it provides access to collateralized credit. In reality, the available collateralized credit to borrowers is determined by his/her income and wealth.

Since earnings increase along the life cycle, it should be the case that poorer and younger households are liquidity constrained, their user cost is higher than the rental price and they prefer renting to owning while they accumulate the down payment needed to buy a house—the findings by Haurin et al. (1996), Linneman et al. (1997), and Quercia et al. (2003) indicate that younger and poorer households are credit constrained in the U.S. economy.

Another key feature of homeownership is that selling housing stock involves substantial transaction costs. This ex-post illiquidity of owner-occupied housing is the

source of most costs to homeownership. In the absence of transaction costs, buying and selling housing stock is not different from buying and selling financial assets. Buying and selling costs affect homeownership in different ways. Higher buying costs delay homeownership over the life cycle since it amounts to a higher down payment. Selling costs discourage young households from becoming owners since they face higher income uncertainty and move more frequently than older households. Selling costs also lower the frequency at which homeowners upgrade or downgrade their houses, as shown by Dıaz and Luengo-Prado (2010). Bajari, Chan, Krueger, and Miller (2010) estimate that, due to substantial transaction costs, most homeowners do not reduce their housing consumption in the short run after a fall in house prices and income, they rather reduce their non housing consumption and deplete their savings. That is, an ex-ante welfare increasing policy may have ex-post welfare decreasing consequences.

The literature on the impact of homeownership on the economic and social behaviors, as well as outcomes (including that of offspring) is extensive, and Dietz and Haurin (2003) provide a review of the micro-level evidence. In general, homeownership is associated to good outcomes (higher success in the labor market, higher education achievement for children, better neighborhood, etc.) but Dietz and Haurin (2003) cast some doubt on some of these results as the standard methodology failed to control for unobserved differences between owners and renters in several cases. The events of recent years, a period of easy credit followed by a downturn coupled with house-price declines, have highlighted some of the dangers of over investing in housing and its broader consequences for the economy.

3 Hypothesis Development

There exists variation in the observed homeownership rates across regions, and many factors affect the propensity to own for an individual. My objective is to investigate factors that contribute to cross-sectional variation in homeownership at the zip code level, with special emphasis on the regional location of the zip code.

3.1 Effect of demographic variables

Age of household head. Age of the household head is typically included in many studies related to tenure choice and homeownership (Gillingham and Hagemann (1983), King (1980), Goodman (1988), Mayo (1981), Lee and Myers (2003), Pickvance (1974) among others). Previous tenure studies have shown that older households are more likely to be homeowners (Eilbott and Binkowski (1985) , Coulson (2002)). To some extent this may reflect the fact that age is a proxy for wealth, and wealthy households are more likely to own homes (Rosen et al. (2017)).

Goodman and Mayer (2018) find that homeownership rate is the lowest for household heads aged 35 years or lesser, increases with age and peaks during retirement after the age of 65. Older people have more working experience and therefore, higher, and more stable income compared to young people (Güriş et al. (2011)). With the expectation of higher wealth of older household head, just like his income, a positive relation is expected between the probability of homeownership and wealth. An older household head has likely accumulated the necessary wealth to overcome downpayment constraint on buying a home (Bourassa (1995)).

Eilbott and Binkowski (1985) show that MSAs with younger populations have lower homeownership rates. Specifically, they find that the percentage of population under 35 has a significantly negative effect on the MSA homeownership. This is also likely due to preference for more inter-regional mobility that comes with renting. Goodman and Mayer (2018) note that areas with high homeownership rates have witnessed lower labor mobility, longer commute times, and lower rates of business formation. Similarly, Lerbs and Oberst (2014) show that higher proportions of young adults is negatively associated with homeownership due to high expected mobility rates (e.g., changing workplaces).

H1a: Zip codes with a higher proportion of relatively younger households have lower homeownership rates.

Type of household. Household type is an important determinant of homeownership (Struyk and Marshall (1974), Gillingham and Hagemann (1983)). A household is composed of one or more people who occupy a housing unit. Non-family house-

holds consist of people who live alone or who share their residence with unrelated individuals.

Goodman and Mayer (2018) and Coulson (2002) find that married couples are much more apt to be homeowners than either those living alone or single householders living with other relatives. Buying a house is the biggest financial decision for a household. Individuals are more comfortable in committing to a mortgage with a spouse than individually or with an unrelated person. Rosen et al. (2017) also find that older, more established married families are more readily able to sustain homeownership. Couples combining their wealth and income would make it easier to own a house compared to singles.

Lee and Myers (2003) argue that since non-family households are mainly composed of single persons or singles with friends, there is no reason to own a home. In fact, they are least likely to own a home among all household types. Larger households, and particularly families with minors, tend to favor owning over renting due the demand for more living space (e.g., gardens) or a higher demand for housing autonomy (Lerbs and Oberst (2014)). Married couples also prefer more settled life compared to singles (Güriş et al. (2011)). As a result, I expect the following hypothesis to hold true,

H1b: Zip codes with a higher proportion of non-family households have lower homeownership rates.

Education level of household head. Education is usually seen to be representative of the lifetime earnings of an individual. The literature indicates a strong association between education and homeownership. Researchers (Painter et al. (2001); Gyourko and Linneman (1996)) note that education may proxy for permanent income, or it could be an indicator for parental resources. Educational attainment is typically positive and significant in models estimating homeownership likelihood (e.g., Coulson (1999); Gyourko and Linneman (1996)). Painter et al. (2001) find that receiving a four year college degree is associated with a 4.1 percentage point higher probability of homeownership among movers. Similarly, Hilber and Liu (2008) find that high school and college completion exhibit a positive impact on homeownership.

Education generally increases both job opportunities and income. In general,

increased earnings potential from additional education makes saving for a down payment, qualifying for a mortgage and affording monthly mortgage payments easier (Rosen et al. (2017)). In fact, compared with not having a high school diploma, a bachelor's degree increased homeownership by 17.5 percentage points for young adults as of 2013, according to a study by Fannie Mae (Myers et al. (2016)). A well-established relationship exists in the housing literature between homeownership and household income and wealth (e.g., Haurin et al. (1995)). Because not all data sources record household wealth, significant attention has focused on down payment constraints (Linneman et al. (1997) among others) that presumably proxy for household wealth and other credit constraints. This literature has drawn consistent conclusions that permanent income is positively related to homeownership and that down payment constraints, either in the form of a lack of assets or in terms of other credit constraints, reduce homeownership.

Education may advance the chances for homeownership in a variety of other ways. Education may have independent effects as a proxy for financial skills, such as knowledge of credit markets, which would increase the chances for homeownership. Grant (2007) shows that income profiles are steeper for better educated households, and hence these households should be more optimistic about their future income. Education may be correlated with factors such as job security and wealth, both of which increase an applicant's ability to obtain a mortgage. Holmes and Horvitz (1994) find that the percentage of adult residents who have at least one year of formal education after high school has a positive and highly significant relationship with conventional mortgages granted. Lusardi (2003) finds that households whose head has a high education have higher savings. Moreover, many households, with low education, have little wealth until late in their life-cycle or start saving very late, making it difficult to do much accumulating (Lusardi (2008)).

According to Birkenmaier and Tyuse (2006), households with less education often have weaker financial education than those with higher education. Collins (2009) finds that mortgage loan applicants with higher rates of college completion are less likely to submit incomplete applications, to withdraw applications before they are reviewed by the lender, and to reject lender-approved loan offers. To the extent

education is correlated with financial capability, these findings suggest loan applicants with lower educational attainment lack financial literacy concerning refinance mortgage application search strategies.

Higher levels of education also facilitate better comprehension of loan options and the solicitation of more options. Hullgren and Söderberg (2013) show that a lower level of education, lower income, lower financial literacy, and trouble handling interest rate influences Swedish consumers to opt for Adjustable-Rate Mortgages (ARMs) which are relatively risky compared to fixed-rate mortgages. Crosby and Taylor (1981) emphasize that better educates and high income consumers have better ability to process information to make a decision and are more adept at computations. Prior studies indicate strong associations among income, education, and performance across a variety of financial capability measures (Hilgert et al. (2003); Lusardi and Mitchelli (2007)).

Additionally, it is likely that higher degrees do equate with higher credit scores (Priceonomics (2017)) due to better decision-making skills (Bernheim et al. (2001)). Using data from the Survey of Consumer Finances (SCF), Ionescu and Simpson (2010) find that people with better credit are more likely to have a bachelor's degree than people with worse credit. By this reasoning, areas with a higher percentage of household-heads holding a bachelors or higher degree would observe a higher homeownership rate due to easier access to mortgage credit. This leads to the next hypothesis,

H1c: Zip codes with a higher proportion of household heads with at least a bachelors' degree have higher homeownership rates.

3.2 Effect of density

The housing industry generally defines multifamily housing as structures with five or more housing units. Housing units in multifamily structures are typically characterized by lesser dwelling area and higher density. While apartments are found both in cities and in the suburbs that surround those cities, they are traditionally associated with urban living. Lerbs and Oberst (2014) argue that urbanized areas are typically

characterized by high proportions of multifamily dwellings in the housing stock, increased levels of land scarcity, and high ownership operating expenses, all of which should impede homeownership.

Multifamily housing structures are also more often associated with rental occupants than owner occupants. Colton et al. (2001) note that multifamily housing tends to be located closer to employment centers, public transportation and neighborhood services. Lack of maintenance, ease of relocation, and convenience in terms of location and transportation are some of the factors which encourage households to occupy rental multifamily housing. Rosen et al. (2017) also point out that renting is more common and buying is less affordable in urban neighborhoods compared to suburban locations.

Coulson (2002) shows that market-based characteristics like price, density, and central city location are most important in explaining regional differences in homeownership. According to Halket et al. (2015), homeownership is correlated with housing structure in England. They suggest that large, detached, expensive, suburban houses are more likely to be owned while smaller, cheaper apartments in urban centers are less likely to be owned. Amior and Halket (2014) also observe that there is striking difference in the homeownership rates of single versus multifamily housing units throughout the U.S. Glaeser and Shapiro (2003) find in their calculations based on The 1998 Survey of Consumer Finances, that 85.5% of people living in single family detached homes are owners, and 85.9% of people living in multifamily units are renters.

Based on this, the proportion of multifamily housing structures in a zip code can proxy for the relative population density or central city location of that zip code. Eilbott and Binkowski (1985) show that heavy concentration of units located in multifamily dwellings in an area might restrain homeownership. High levels of population density may, therefore, be associated with lower ownership rates. I predict the following hypothesis to hold true,

H2: Zip codes with a relatively higher proportion of multifamily housing structures have lower homeownership rates.

3.3 Effect of housing market variables

Regional housing prices react uniformly to certain national economic factors, such as mortgage rates. On the other hand, local factors such as population shifts, employment and income trends often have unique impact on housing prices. Reichert (1990) rejects the hypothesis of a single national housing market in favor of one that allows for broad national trends to be superimposed upon unique regional markets. His study finds that during the period 1975-1987, housing prices in the Northeast appear to be most sensitive to changes in construction costs while those in the West were most sensitive to permanent income.

Baffoe-Bonnie (1998) finds that economic variables have a different impact on the dynamic behavior of housing prices and the number of houses sold in different regions at different time periods. The results of this study suggest that housing prices and houses sold do respond to economic fundamentals. These economic conditions may vary from one region to another. Differences in structural conditions in the various regions may account for regional differences in houses sold and housing prices (for example, supply elasticities are relatively low on the West and East coasts but high in the South and Southwest parts of the United States). Glaeser (2007) points that the expensive places in the U.S. have low development and the places in the U.S. with abundant development have low prices. This indicates differences in housing supply across regions due to land use regulations, construction costs or land density. Gyourko and Saiz (2006) estimate that one-fifth of the variation in housing prices across space can be attributed to differences in the cost of building homes.

Glaeser (2007) also notes that a positive correlation between housing values and median household incomes across MSAs could imply causation in either direction. It might mean that some places are intrinsically more productive than others and therefore have high wages which then cause high prices. Alternatively, it might mean that some places are intrinsically more expensive than others and those high prices then ensure the need to pay high wages. Ortalo-Magné and Prat (2014) investigate mechanisms that generate urban growth restrictions that are critical to housing market dynamics. They hypothesise that new housing construction in an

area implies an immediate drop in the price of housing and reduces the future housing rents. Based on this, I expect the following hypothesis to hold true,

H3: After controlling for regional fixed effects, the conditional effects of housing market variables are different for zip codes located in different regions.

Price-Rent Ratio. The relationship between house price and rent is an important concept in the analysis of housing markets. The price-rent ratio in the housing market is like the price-to-earnings ratio for stocks (Leamer (2002), Himmelberg et al. (2005)). This metric is intended to reflect the relative cost of owning versus renting. Goswami et al. (2014) treat housing as an asset that entitles the owner to a series of future rent cash flows. Price-rent ratios offer important housing market insights. Variations across time and space in price-rent ratios can provide information about local real estate markets, including future market expectations and speculative market behavior.

Current price-rent ratio reflects households' expectations about future rent growth and future housing prices (Kishor and Morley (2015)). An upward surprise in the price–rent ratio today must correspond to news that future housing returns will be higher or to a downward revision in expected rent growth. Peng and Yang (2015) derive the minimum price-rent ratio below which a household will find buying more attractive than renting. As a result, if the current market price-rent ratio is above this equilibrium, the household will self-select to be renter. A high price-rent ratio implies that either housing prices are expected to fall or rents are expected to rise. Both of these market movements will motivate households to reduce current ownership.

This effect can be termed as the long-run steady-state effect. Homeownership is a long-term decision based as much upon anticipated future needs as on present needs. A household who plans to remain in its home over a long period is more likely to own; a household who plans to remain in its home over a short period is more likely to rent. Different housing markets have different equilibrium or steady-state value of price-rent ratio. If there is any deviation from this long-run equilibrium value, the price-rent ratio should self-correct. Gelain and Lansing (2014) demonstrate that an upward shift in households' preference for housing raises the mean price–rent

ratio and lowers the average realized return on housing. Since homeownership rate is the cumulative result of households' house price expectations in the future, a higher price-rent ratio in an area is expected to lead to a lower aggregate homeownership rate in that area. A unified national housing market would indicate a common steady state of price-rent ratio and hence, zip codes with a lower value having a higher homeownership rate.

This paper focuses on variation across space and the impact of price-rent ratio on homeownership. Himmelberg et al. (2005) argue that differences in appreciation rates and local taxes explain disparities in price-rent ratios across markets and find that prices-rent ratios remain fairly consistent over time. Begley et al. (2019) also show that the price-rent ratio moves similarly across all property types, but varies significantly by geography. Himmelberg et al. (2005) assert that one cannot draw conclusions about house prices by comparing cities: price-to-income and price-to-rent ratios that would be considered "high" for one city may be typical for another.

Focusing specifically on spatial variation in price-rent ratios, Capozza and Seguin (1996) find that price-rent ratios are predictive of future price appreciation within metropolitan areas. Sinai and Souleles (2005) find statistical evidence using data from 44 metropolitan areas that the cities with the highest price-to-rent ratios also have the highest expected growth rates of prices and rents. Later work by Gallin (2008) also confirms that price-rent ratios are predictive of future prices. Thus, higher price-rent ratio in an area directly leads to higher homeownership due to households' expectations of higher future growth rate.

This can be called a short-term effect within region. Prices within a region are more likely to be correlated compared to housing prices across the country. Ortalo-Magné and Rady (2002) predict that all else being equal, the advantage of ownership increases with the covariance of the prices of the different properties that a household may consider for present and future housing consumption. To the extent that households prefer to stay in one region due to reasons like employment prospects and community ties, small differences in costs of renting and owning may lead to none or positive effects on homeownership within a region (Pickvance (1974)).

There is another channel through which price to rent ratio can be expected to

positively correlate to homeownership. Cronin and McQuinn (2016) examine the implications of macroprudential policy and, in particular, loan-to-value (LTV) limits on the house price to rent ratio. They find that a policy, which lowers the effective LTV ratio will lead to a permanent decline in the house price to rent ratio. Higher LTV ratios in an area lead to higher homeownership by relaxing households' down-payment constraint. Lamont and Stein (1997) empirically show in cities where LTV ratios are higher, price-rent ratios are also higher.

Jurgilas and Lansing (2012) develop a theoretical model where easier lending standards and lower mortgage transaction costs contribute to a substantial rise in house prices relative to rents by reducing households' perception of economic risk. Lower risk perception induces households to accept a lower rate of return on the purchase of risky assets such as houses. A lower expected return leads to an increase in the area's fundamental price-rent ratio.

Sinai and Souleles (2005) emphasize the role of homeownership as a hedge against fluctuations in rents. They find that rent risk (volatility of rent growth) leads to higher housing prices relative to rents. However, the relationship between house price and rent varies across zip codes due to the composition of housing structures. As Hattapoglu and Hoxha (2014) indicate that failing to control for different types of houses poses a problem. Previous research has provided ambiguous results which may have been caused by comparing prices of single family houses with rents, where the latter is dominated by apartments.

Thus, I investigate the following hypothesis,

H3a: Zip codes with a higher ratio of median home value to median rent have lower homeownership rates, conditional on everything else. However, a non-negative relationship between homeownership rate and price-rent ratio is to be expected when regional differences are taken into account. If each region can be considered a different housing market, they are also expected to have different equilibrium price-rent ratio.

Price-Income Ratio. Price-to-income ratio is another key measure used to assess whether housing prices are "too high" (Himmelberg et al. (2005)). Unlike the price-to-rent ratio, which measures the relative cost of owning and renting, the price-to-income ratio provides a measure of local housing costs relative to the local ability

to pay. Nissan and Payne (2013) find that regional housing markets may be driven by local demand and supply factors, and that the response of regional housing prices to business cycle fluctuations may be asymmetric across regions.

Income (current and expected lifetime) directly impacts on a household's ability to purchase and make housing payments. Households with insufficient wealth or income (relative to their preferred housing consumption and local house prices) are unable to become homeowners due to constrained access to mortgages (Acolin et al. (2016)). Adelino et al. (2018) find that higher income households perceive housing as significantly safer than low-income individuals, perhaps because they find it easier to navigate the risks of the housing market.

The most widely used and cited indicator of homeownership affordability is the median house price to median income ratio, due to its simplicity and ease of understanding (Phang (2009)). However, this ratio fails to incorporate many factors that affect the affordability of housing. The main factor not directly considered by this ratio is the prevailing interest rate. Since the majority of house purchases involve a loan, the interest rate is an important influence on people's ability to pay. Assuming that interest rates are the same for all regions in a country, the difference lies in house prices and incomes. Demand for housing is a function of many factors, but mainly the desirability of each location (Robinson et al. (2006)). Based on this, I investigate the following hypothesis,

H3b: Zip codes with a higher ratio of median home value to median household income have lower homeownership rates. Moreover, the regression coefficient of homeownership on price-income ratio is expected to be different across regions.

4 Data

4.1 Census Regions

The U.S. Census Bureau has defined a system of state groupings called 'regions' (appendix B) based on contiguous counties with the primary purpose of classification for objective statistical analysis. Regions combine all the nation's counties

and statistically equivalent entities into a more concise set of general-purpose areas. The four U.S. regions have very distinct geographical, climatic, demographic, and economic characteristics and hence, characteristically different housing markets. Socioeconomic heterogeneity was the principal criterion for grouping states into regions.

Table 1 presents the descriptive statistics for demographic and economic variables that affect homeownership outcomes in an area. West and Northeast regions have a higher proportion of multifamily housing structures, higher proportion of more educated household-heads, higher rents and home values, and higher incomes compared to Midwest and South. While these may imply similarities between West and Northeast, they are very different on the age-profile of household heads. West has the highest proportion of relatively younger households and Northeast has the highest proportion of older households.

As a result of these socio-demographic and economic differences, there is variation in the homeownership rate across regions. Figure 2 shows the quarterly homeownership rates for nationwide and the four regions for 1990:Q1 - 2017:Q1 from the Federal Reserve Economic Data. For the time period 2007-2011, West had the lowest homeownership rate (61.28%) and Midwest had the highest (70.32%). Midwest and South regions have always had higher homeownership rates than the other regions and higher than the national average. During this period, homeownership reached its peak in 2005:Q1 and has been declining since then, the decline being most steep in the West.

4.2 Sample Data

The U.S. Census Bureau reports the median housing prices and other characteristics as part of the Housing Vacancy Survey (HVS), Current Population Survey (CPS) and the American Community Survey (ACS). While Census data has the broadest coverage, it reports only median prices for owner occupied homes. For this study, I use the aggregate 5-year data from the 2011 American Community Survey (ACS) at the zip code level. These contain 'period' estimates for data collected over the time period 2007-2011. The primary advantage of using multiyear estimates is the

Table 1: **Region Descriptive Statistics, ACS 2007-2011**

Variable	U.S.	West	Midwest	Northeast	South
Homeownership Rate (%)	66.13	61.28	70.32	63.56	67.74
Household-heads aged less than 35 years (%)	20.40	21.50	20.70	17.40	20.70
Household-heads aged 65 years and over (%)	21.50	20.00	21.90	22.80	21.30
Median household income ($ '000)	52.8	57.4	51.1	59.0	48.8
Median home value ($ '000)	186.2	296.6	150.3	261.1	154.0
Median gross rent ($)	871	1,034	734	972	817
Non-family households (%)	33.30	32.70	34.30	34.50	32.50
Housing structures with more than five units (%)	17.62	19.70	14.80	21.30	16.30
Household-heads with Bachelors' degree or higher (%)	25.72	26.73	24.57	29.88	23.73

Source: 2007-11 American Community Survey, U.S. Census Bureau
Note 1: Numbers represent officially reported aggregate estimates for the U.S. and each region.
Note 2: The 2011 ACS sampled approximately 3.3 million housing unit addresses.

Table 2: **Sample Descriptive Statistics**

Variable	U.S.	West	Midwest	Northeast	South
$HOR(\%)$	74.40	67.96	78.18	75.27	73.38
	(17.38)	(19.49)	(14.44)	(18.59)	(17.03)
$hhUnder35yrs(\%)$	16.45	17.26	16.84	14.32	16.97
	(10.80)	(12.42)	(9.73)	(9.91)	(11.30)
$hhOver65yrs(\%)$	24.91	23.45	25.67	24.87	24.95
	(12.01)	(13.90)	(10.76)	(11.29)	(12.51)
$priceRentRatio$	18.75	25.87	16.05	21.52	15.82
	(11.89)	(18.36)	(7.31)	(11.32)	(9.26)
$priceIncomeRatio$	3.45	5.43	2.47	3.93	3.08
	(4.35)	(7.07)	(1.59)	(2.59)	(4.88)
$propNonFamily(\%)$	31.07	31.80	30.70	31.69	30.66
	(13.25)	(15.77)	(11.71)	(13.00)	(13.37)
$moreThan5Units(\%)$	7.62	9.74	5.75	9.30	7.36
	(13.74)	(15.66)	(10.58)	(15.96)	(13.74)
$bachelors(\%)$	20.35	22.96	17.57	29.08	18.12
	(14.86)	(16.40)	(11.80)	(16.85)	(14.20)

Source: 2007-11 American Community Survey, U.S. Census Bureau

Note 1: Sample consists of 29,851 5-digit zip codes. Of these, 5899, 9894, 8865, 5193 observations belong to the Northeast, Midwest, South and West regions, respectively.

Note 2: Numbers represent the mean values for each variable. Mean values for U.S. are computed over the entire sample. Numbers in parentheses represent standard deviations.

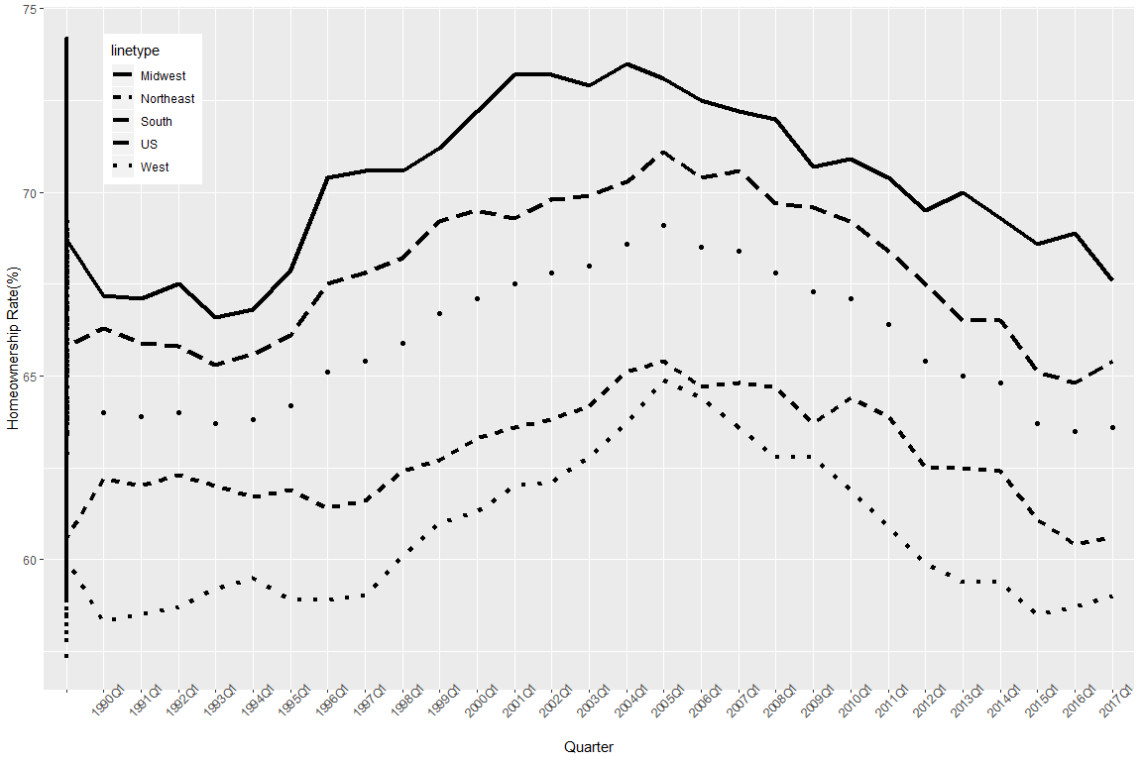


Figure 2: **Homeownership Rate for the U.S. and Regions , 1990 - 2017**

Source: https://fred.stlouisfed.org/

increased statistical reliability of the data for less populated areas and small population subgroups. ACS 5-year estimates were first released for the period 2005-2009. The 5-year ACS is unique in its ability to produce estimates on 35 housing variables amongst other key characteristics at extremely low geographic levels.

The unit of observation in this study is a Census-developed proxy for zip codes: Zip Code Tabulation Areas (ZCTAs). "ZCTAs are generalized area representations of U.S. Postal Service (USPS) zip code service areas. Simply put, each one is built by aggregating the Census 2000 blocks, whose addresses use a given zip code, into a ZCTA which gets that zip code assigned as its ZCTA code. They represent the majority USPS five-digit zip code found in a given area. For those areas where it is difficult to determine the prevailing five-digit zip Code, the higher-level three-digit

zip code is used for the ZCTA code."[1]

I selected ZCTAs that carry the full five digit zip code and assigned region membership to these using the state Federal Information Processing Standards (FIPS) codes. Following the U.S. Census Bureau data documentation, I exclude ZCTAs that either had no sample observations or too few sample observations to compute an estimate. Additionally, if the median in a ZCTA falls in the lowest interval of the distribution, the monthly rent for that ZCTA has been winsorized at \$100 and the home value has been winsorized at \$10,000. Similarly, if the median in a ZCTA falls in the uppermost interval of the distribution, the monthly rent for that ZCTA has been winsorized at \$2000 and the home value has been winsorized at \$1,000,000.

Of the total of 29851 observations, 5899, 9894, 8865, 5193 observations belong to the Northeast, Midwest, South and West regions, respectively. These comprise of both urban and rural zip codes. Table 2 reports the mean values of all explanatory factors for the U.S. and the four regions. Looking at simple averages of the observations, West has the lowest homeownership rate followed by South, Northeast and Midwest. West also has the highest proportion of relatively younger households, the highest price-rent ratio, highest price-income ratio and highest proportion of multifamily housing structures on average.

The U.S. Census Bureau collects information on these variables at the household level and aggregates those over different geographies. Such estimates are reported in 1 for the U.S. and four region. These must not be confused with the aggregate estimates from Table 2, which are in contrast with these. In this study, I use the aggregates reported at the zip code level and use the simple average of all those belonging to the same region, solely for the purpose of discussion.

5 Econometric Model

I use demographic and household characteristics for each ZCTA and three dummy variables to explain variation in homeownership rates across regions. Three dummy

[1] https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html

variables are used in a linear regression model with intercept to avoid dummy variable trap with four regions. I choose 'West' as the omitted category since it has the lowest homeownership rate throughout my time period of interest, 2007-2011 (Table 1).

The dependent variable is the percentage of occupied housing units that were owner occupied in the period 2007-2011 in each of the ZCTAs. The independent variables are:

x_1 - *hhUnder35yrs.* This represents the percentage of household heads in a ZCTA that are younger than 35 years of age.

x_2 - *hhOver65yrs.* This represents the percentage of household heads aged 65 years or more.

x_3 - *priceRentRatio.* This variable is computed as the ratio of median home value in a ZCTA to median annual rent.

x_4 - *priceIncRatio.* This variable is computed as the ratio of median home value in a ZCTA to median household income.

x_5 - *propNonFamily.* This measures the percentage of non-family households in a ZCTA.

x_6 - *moreThan5Units.* This measures the percentage of housing structures with five or more units.

x_7 - *bachelors.* This is the percentage of household heads with a Bachelors' degree or higher.

d_1, d_2, d_3 - *Census regional location of the ZCTA, where*
$d_1 = 1$ if ZCTA is in Midwest, 0 otherwise.
$d_2 = 1$ if ZCTA is in Northeast, 0 otherwise.
$d_3 = 1$ if ZCTA is in South, 0 otherwise.
$d_1 = d_2 = d_3 = 0$ if ZCTA is in West.

I use the following regression model to explain variation in homeownership rates,

$$\begin{aligned} HOR_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i} \\ + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \gamma_3 d_{3i} \\ + \beta_{31} x_{3i} d_{1i} + \beta_{32} x_{3i} d_{2i} + \beta_{33} x_{3i} d_{3i} \\ + \beta_{41} x_{4i} d_{1i} + \beta_{42} x_{4i} d_{2i} + \beta_{43} x_{4i} d_{3i} \end{aligned} \tag{1}$$

where,
i is the identifier for a ZCTA,
β_j is the slope coefficient for variable j for West, and
$(\beta_j + \beta_{jk})$ is the slope coefficient for variable j for,
Midwest if $k = 1$,
Northeast if $k = 2$,
South if $k = 3$.
γ_k is the slope coefficient for dummy variable k.

While this OLS model is useful as a reference, it has a few methodological drawbacks. First, each unit of observation (ZCTA) represents a region located in space. This makes spatial dependence between the observations a likely scenario due to possible linkages. In the presence of spatial dependence, the OLS estimator is known to be no longer the best linear unbiased estimator. A second potential source of bias is the existence of unobserved regional characteristics that exert an influence on regional homeownership rates. For instance, the share of owner-occupiers in a given area may be affected by unobserved preferences for homeownership, e.g. regional traditions. If any of the omitted factors correlate with the explanatory variables, their influence may wrongly be attributed to the included covariates. Using cross-sectional OLS - i.e., in the absence of a panel - it is not possible to control directly for unobserved regional characteristics.

Next, I evaluate the following linear models,

Model A:

$$HOR_i = \gamma_0 + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \gamma_3 d_{3i}$$

Model B:

$$HOR_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \gamma_3 d_{3i}$$

Model A is a baseline model with the homeownership rates regressed only on the regional dummy variables with *West* as the omitted category. Model B includes all the covariates without any interaction terms so as to focus only on the regional variation.

To quantify the variability into a single measure, I define the sum of squared differences (SSD) as the sum of squares of pairwise differences between regression coefficients on the dummy variables,

$$SSD = \sum_{i=0}^{2} \sum_{j=i+1}^{3} (d_i - d_j)^2$$

where d_0, d_1, d_2 and d_3 indicate West, Midwest, Northeast and South respectively. The reduction in SSD for *Model B* relative to SSD for *Model A* can be interpreted as the regional variation that can be accounted for by my included explanatory variables.

5.1 Model Selection

Based on economic theory and existing research, I identify seven demographic and market variables that affect the homeownership rate in a ZCTA besides the regional location of the ZCTA. I adopt a general-to-simple[2], or backward step wise reduction model to achieve a simpler specification and avoid overfitting, if any.

I run regressions on the full unrestricted model and also iteratively by dropping one explanatory variable at a time. Next, I compare the explanatory power of each

[2] Greene (2003)

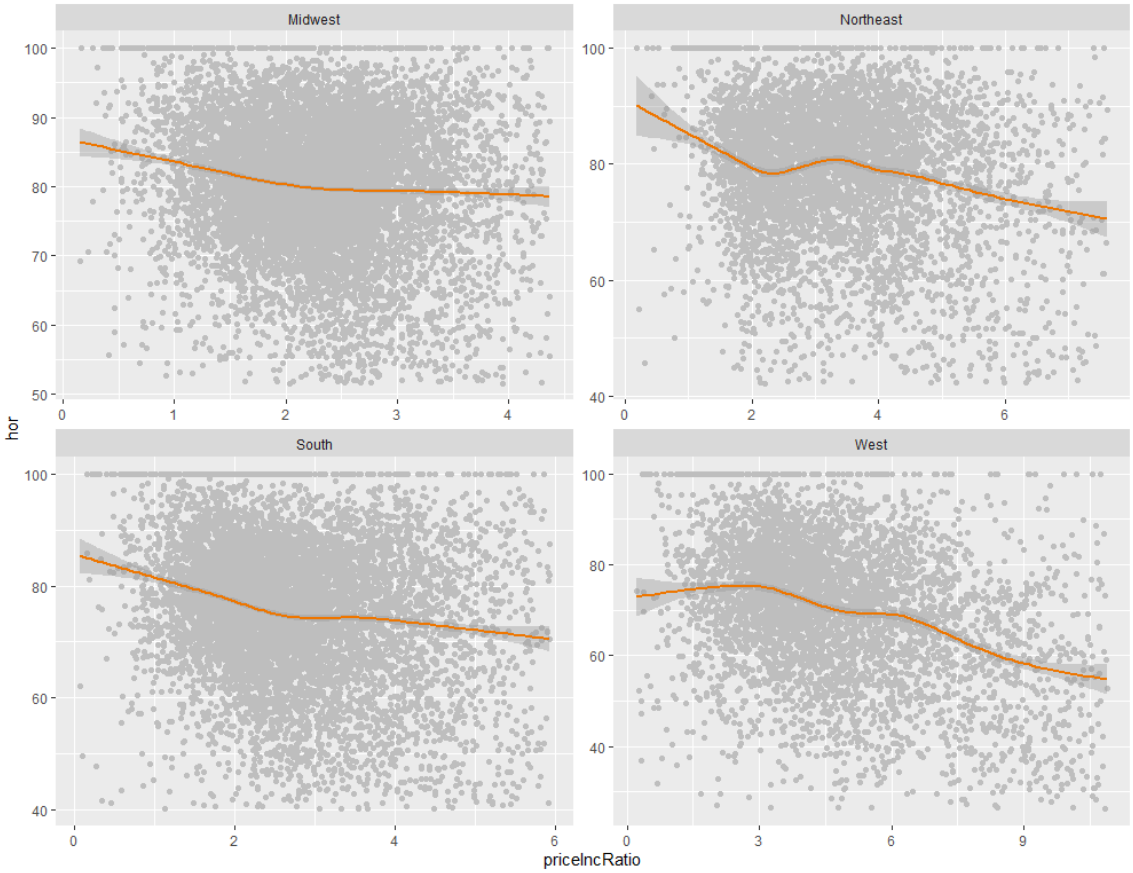


Figure 3: **Price-Income Ratio across Regions**

Source: 2007-2011 American Community Survey
Note 1: The figure shows the relationship between price-income ratio and the homeownership rate for each region. The relationship is always negative but the line of best fit also has different slopes for different regions.
Note 2: Outliers for both price-income ratio and homeownership rate have been dropped for each region. Specifically, only the following range of values are plotted in the figure,

West: $priceIncRatio < 10.89 \quad \& \quad HOR > 41.39$
Midwest: $priceIncRatio < 4.37 \quad \& \quad HOR > 51.53$
Northeast: $priceIncRatio < 7.64 \quad \& \quad HOR > 42.20$
South: $priceIncRatio < 5.94 \quad \& \quad HOR > 40.09$

Note 3: The following limits were used for range of values after dropping outliers,

$$(Q_1 - 1.5 * IQR, \quad Q_3 + 1.5 * IQR)$$

where,
$IQR = Q_3 - Q_1$,
Q_1 and Q_3 are the first and third quantiles, respectively.

of the reduced models to the unrestricted model. If the adjusted-R^2 of a restricted model obtained by dropping a variable falls by more than 3% compared to the unrestricted model, then I proceed with retaining that variable in my model. The choice of 3% threshold is arbitrary.

Adjusted-R^2 has been suggested as a measure of goodness of fit that appropriately penalizes the loss of degrees of freedom that result from adding variables to the model. Two alternative fit measures that are commonly considered in model selection are the Akaike Information Criterion (AIC) and the Schwarz or Bayesian Information Criterion (BIC). Both measures improve as adjusted-R^2 increases, but, everything else constant, degrade as the model size increases. Unlike adjusted-R^2, a lower AIC or BIC value indicates a better fit.

I follow this approach for two model specifications to ensure consistency. Table 3 reports the results of eight regression models starting from the unrestricted model in the first column and then dropping one variable at a time. This specification includes only the quantitative explanatory variables and omits any regional factor. Similarly, Table 4 reports results for the second specification where regional dummy variables as well as interaction terms with price-rent ratio are included in the models. Finally from Table 5, I select the model with the highest adjusted-R^2 and lowest AIC and BIC. This specification includes additional terms for interaction between region and price-income ratio but reports results only for the selected models for comparison.

Table 3 reports the results of eight regression models. Model 1 in the first column includes all the variables. Based on the highest adjusted-R^2 and the lowest AIC and BIC, Model 1 has the best explanatory power. One variable is dropped at a time in Model 2 – 8. The adjusted-R^2 reduces most heavily from 64% to 51.8% (by 19.1% or 12.2 percentage points) by omitting the variable, $moreThan5Units$ in Model 3. Similarly, omission of $bachelors$, $propNonFamily$, $priceRentRatio$ and $hhUnder35years$ also lead to substantial loss of explanatory power. In fact, $hhOver65years$ and $priceIncRatio$ are the only variables which when dropped individually do not lead to a decline in adjusted-R^2 of more than 3%.

I repeat this exercise and report results in Table 4, except that I also include locational variables. Due to the strong presence of regional variation in homeowner-

Table 3: **Model Selection I**

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
x_0 - intercept	88.3094***	93.7026***	99.7083***	84.2078***	88.2289***	87.1196***	91.8889***	77.2432***
x_1 - hhUnder35yrs	-0.4409***	-0.5191***	-0.7141***	-0.5029***	-0.4424***	-0.4218***	-0.5119***	—
x_2 - hhOver65yrs	0.1321***	0.0809***	0.1974***	0.0239**	0.1308***	0.1150***	—	0.3253***
x_3 - priceRentRatio	-0.0649***	0.0430***	-0.0805***	-0.0569***	-0.1126***	—	-0.0672***	-0.0268***
x_4 - priceIncRatio	-0.4140***	-0.4497***	-0.7725***	-0.4656***	—	-0.2437***	-0.4111***	-0.4230***
x_5 - propNonFamily	-0.2769***	-0.2843***	-0.5485***	—	-0.2850***	-0.2126***	-0.2404***	-0.3340***
x_6 - moreThan5Units	-0.5688***	-0.4638***	—	-0.6808	-0.5921***	-0.6351***	-0.5779***	-0.6724***
x_7 - bachelors	0.2577***	—	0.1099***	0.2620***	0.2610***	0.1985***	0.2476***	0.2995***
adj-R^2	0.6400	0.6007	0.5178	0.6154	0.6342	0.5894	0.6360	0.6040
AIC	189252	191950	196862	190967	189665	218398	189534	191728
BIC	189326	192015	196927	191032	189730	218465	189600	191794

Note 1: The table shows the results of eight regression models investigating the relationship between various factors and homeownership, without any regional effects.
Note 2: Model 1 includes all the explanatory variables. Models 2-8 exclude one variable each.
Note 3: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 4: **Model Selection II**

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
x_0 - intercept	84.2121***	90.2726***	95.2874***	80.6970***	83.7687***	82.5785***	87.0844***	73.1187***
x_1 - hhUnder35yrs	-0.4530***	-0.5146***	-0.7298***	-0.5149***	-0.4549***	-0.4348***	-0.5123***	—
x_2 - hhOver65yrs	0.1095***	0.0623***	0.1728***	0.0008	0.1077***	0.0991***	—	0.3062***
x_3 - priceRentRatio	-0.0640***	-0.0052	-0.0595***	-0.0761***	-0.0849***	—	-0.0662***	-0.0352***
x_4 - priceIncRatio	-0.2435***	-0.2971***	-0.5921***	-0.2993***	—	-0.1234***	-0.2380***	-0.2630***
x_5 - propNonFamily	-0.2798***	-0.2830***	-0.5515***	—	-0.2847***	-0.2177***	-0.2496***	-0.3368***
x_6 - moreThan5Units	-0.5644***	-0.4691***	—	-0.6780***	-0.5774***	-0.6256***	-0.5718***	-0.6712***
x_7 - bachelors	0.2565***	—	0.1174***	0.2585***	0.2594***	0.2097***	0.2485***	0.2912***
d_1 - mw	3.9130***	2.2745***	5.1443***	2.7334***	4.2924***	7.4708***	4.1085***	3.4946***
d_2 - ne	5.5958***	4.2602***	6.8589***	4.1343***	6.0333***	4.5329***	5.4745***	5.9417***
d_3 - so	2.2656***	0.2693	1.9064***	1.7940***	2.8196***	4.0557***	2.3191***	2.0524***
x_3d_1 - priceRentRatio:mw	0.2107***	0.2804***	0.1513***	0.2634***	0.2140***	—	0.2090***	0.2208***
x_3d_2 - priceRentRatio:ne	-0.0707***	0.0304*	-0.1575***	-0.0171	-0.0809***	—	-0.0654***	-0.0353*
x_3d_3 - priceRentRatio:so	0.0856***	0.1829***	0.0879***	0.1061***	0.0701***	—	0.0898***	0.0951***
adj-R^2	0.6704	0.6333	0.5513	0.6457	0.6685	0.6139	0.6677	0.6335
AIC	186954	189732	194994	188835	187103	216605	187164	189717
BIC	187077	189846	195108	188950	187217	216697	187279	189831

Note 1: The table shows the results of eight regression models investigating the relationship between various factors and homeownership, along with regional dummy variables and interaction terms with price-rent ratio.
Note 2: Model 1 includes all the explanatory variables. Models 2-8 exclude one variable each.
Note 3: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

ship rates, I don't exclude the regional dummies from any of the candidate models in Table 4. I find that the results align very well to those in Table 3. The adjusted-R^2 reduces most heavily from 67% to 55% (by 17.8% or about 12 percentage points) by omitting the variable, $moreThan5Units$ in Model 3. Similarly, omission of $bachelors$, $propNonFamily$, $priceRentRatio$ and $hhUnder35years$ also lead to substantial loss of explanatory power. Once again, $hhOver65years$ and $priceIncRatio$ are the only variables which when dropped individually do not lead to a decline in adjusted-R^2 of more than 3%.

Above discussion indicates that conditional on other variables being included in the model, at least one of $hhOver65years$ and $priceIncRatio$ may be dropped from the final model specification. However, it is well established in the housing research that the price-to-income ratio provides an adequate indicator of homeownership affordability in regional housing markets (Lerbs and Oberst (2014)). While on the other hand, it is likely that the presence of another age-related variable, $hhUnder35yrs$ in the regression equation suppresses the explanatory power of $hhOver65years$. These two are also likely to be correlated. Hence, I conclude that it is reasonable to exclude $hhOver65years$ from the regression model and include all other explanatory variables.

Surprisingly, omission of $priceIncRatio$ decreases the explanatory power of the unrestricted model only marginally (by 0.9% in Table 3 and by 0.3% in Table 4). It turns out that $priceIncRatio$ impacts the homeownership rate very differently across different regions. Figure 3 shows the relationship between price-income ratio and the homeownership rate for each region. The relationship is always negative but the line of best fit also has different slopes for different regions. For instance for the Midwest, homeownership rate stays around 80% for price-income ratio greater than 2. They seem to be linearly inversely related when price-income ratio is less than 2. While for Northeast and West, the relationship, though negative, is non-linear.

In order to account for the different nature of relationship between homeownership and price-income ratio across regions, I also include interaction of $priceIncRatio$ with regional dummy variables. Table 5 reports results for four regression specifications. All four models exclude $hhOver65yrs$ from the equation. Model 1 from

Table 5 is the same as Model 7 from Table 4. Model 2 also excludes *priceIncRatio* in addition to $hhOver65yrs$. Consistent with earlier observation, the adjusted-R^2 declines marginally from 66.77% to 66.59%. Model 3 omits the interaction terms for *priceRentRatio* but includes those with *priceIncRatio*. Adjusted-R^2 increases by approximately 1 percentage points relative to Model 1.

Model 4 omits only $hhOver65yrs$ and includes interaction terms with both *priceRentRatio* and *priceIncRatio*. This model has the highest adjusted-R^2 and the lowest AIC and BIC, among all four models. *priceIncRatio* plays a key role in explaining the cross-sectional variation in homeownership rates when regional heterogeneity is accounted for by including interaction terms with regional dummy variables. Therefore, I select the specification in Model 4 from Table 5 as the best model for investigating my hypotheses.

6 Results

Results of the best econometric specification are reported in Model 4 in Table 5. All the demographic variables i.e. $hhUnder35yrs$, $propNonFamily$ and *bachelors*, and $moreThan5Units$ have the expected signs and high statistical significance. The intercept x_0 represents the mean conditional homeownership rate for the West region. All the dummy variables d_1, d_2, d_3 and corresponding interaction terms are interpreted relative to the West region since that is the omitted category.

A striking observation from Table 5 is the instability of the dummy variable for the South across the four models. In Models 1 and 2 which do not include the interaction terms *priceIncRatio:region*, the *so* dummy is equal to 2.3191 and 2.8599, respectively and is significant at 0.001 level of significance. However, it is equal to only 0.6057 in magnitude and becomes less significant in Model 3. This happens when the interaction terms *priceRentRatio:region* are dropped and the terms *priceIncRatio* are added to the model. In the best model specification (Model 4), where the *priceRentRatio:region* terms are also included, the *so* dummy becomes statistically insignificant, i.e., not significantly different from zero.

This occurrence can perhaps be explained by the presence of relatively large intra-

Table 5: **Model Selection III**

Variable	Model 1	Model 2	Model 3	Model 4
x_0 - intercept	87.0844***	86.6057***	89.1007***	89.1880***
x_1 - hhUnder35yrs	-0.5123***	-0.5132***	-0.5304***	-0.5277***
x_2 - hhOver65yrs	—	—	—	—
x_3 - priceRentRatio	-0.0662***	-0.0866***	0.0357***	-0.0018
x_4 - priceIncRatio	-0.2380***	—	-1.1274***	-1.0158***
x_5 - propNonFamily	-0.2496***	-0.2549***	-0.2531***	-0.2438***
x_6 - moreThan5Units	-0.5718***	-0.5843***	-0.5205***	-0.5183***
x_7 - bachelors	0.2485***	0.2515***	0.2450***	0.2421***
d_1 - mw	4.1085***	4.4764***	5.6492***	3.9933***
d_2 - ne	5.4745***	5.9042***	5.7213***	5.0079***
d_3 - so	2.3191***	2.8599***	-0.6057*	0.0133
x_3d_1 - priceRentRatio:mw	0.2090***	0.2122***	—	0.2517***
x_3d_2 - priceRentRatio:ne	-0.0654***	-0.0755***	—	0.0787***
x_3d_3 - priceRentRatio:so	0.0898***	0.0745***	—	-0.0327*
x_4d_1 - priceIncRatio:mw	—	—	0.2134*	-0.7706***
x_4d_2 - priceIncRatio:ne	—	—	-0.6150***	-0.8637***
x_4d_3 - priceIncRatio:so	—	—	1.1494***	1.0741***
adj-R^2	0.6677	0.6659	0.6781	0.6814
AIC	187164	187304	186335	186074
BIC	187279	187411	186450	186214

Note 1: The table shows the results of four regression models investigating the relationship between various explanatory factors and homeownership rate in a ZCTA.
Note 2: Model 1 is the same as Model 7 from table 4.
Note 3: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

regional heterogeneity in housing prices and greater dispersion in household incomes across Southern cities. The South consists of 16 states and 1421 counties (out of the total 3142 counties in the U.S. as of 2013). States with the largest (Texas with 50% drop) and the smallest (Oklahoma with 6% drop) decline in home prices during 2007-2011 co-exist in the South[3]. Additionally, the relative variability as measured by coefficient of variation[4] (CV) is also the highest in the South among all four regions for both the median household income as well as the median house prices. This also leads to the highest CV for price-income ratio in the South.

The net effects for each variable are reported in Table 6. Column 1 (β_i) reports regression coefficients and represents percentage point change in homeownership rate by a 1 unit change in the right-hand side variable. Column 2 (σ_x) reports the standard deviations of respective variables. Column 3, (β_i^*) contains the standardized coefficients. β_i^* represents the change in left-hand side variable in percentage points when the value of i^{th} variable increases by one standard deviation.

6.1 Analysis of Hypotheses 1

$hhUnder35yrs$ is the proportion of householders aged 35 years or less in a ZCTA and is representative of the proportion of relatively younger households in the area. From Model 4 in Table 5, the coefficient on $hhUnder35yrs$ is negative and statistically significant at 0.001 level of significance. This is consistent with hypothesis H1a. Conditional on other included variables, a 1 unit decrease in the proportion of relatively younger households in a ZCTA is associated with an increase of 0.53 percentage points in the homeownership rate in the ZCTA. This conditional effect holds irrespective of the regional location of the ZCTA. From Table 6, the standard deviation of the distribution of $hhUnder35yrs$ is 10.80. Hence, in standardized terms, the effect of 1 standard deviation increase in the proportion of younger households in a ZCTA is associated with an increase of approximately 5.7 percentage points in the homeownership rate.

[3] https://www.corelogic.com/downloadable-docs/corelogic-peak-totrough-final-030118.pdf

[4] Appendix C

Table 6: **Conditional Effects on HOR**

Variable		(1) β_i	(2) σ_i	(3) β_i^*
$hhUnder35yrs$		-0.5277	10.80	-5.6992
$priceRentRatio$	$West$	-0.0018	18.36	-0.0330
	$Midwest$	0.2499	7.31	1.8268
	$Northeast$	0.0769	11.32	0.8705
	$South$	-0.0345	9.26	-0.3195
$priceIncRatio$	$West$	-1.0158	7.07	-7.1817
	$Midwest$	-1.7864	1.59	-2.8404
	$Northeast$	-1.8795	2.59	-4.8679
	$South$	0.0583	4.88	0.2845
$propNonFamily$		-0.2438	13.25	-3.2304
$moreThan5Units$		-0.5183	13.74	-7.1214
$bachelors$		0.2421	22.68	5.4908

Note: The table shows the values of region-specific standardized and unstandardized regression coefficients. β_i represents percentage point change in homeownership rate for a 1 unit change in the value of i^{th} variable. β_i^* represents percentage point change in homeownership rate for a 1 standard deviation change in the value of i^{th} variable. σ_i is the standard deviation of i^{th} variable.

$$\beta_i^* = \beta_i * \sigma_i$$

propNonFamily is the proportion of non-family households in a ZCTA. Consistent with hypothesis H1b, I expect the regression coefficient on this variable to be negative and statistically significant. This can be seen from Model 4 in Table 5. Conditional on all other included variables in the model, a 1 unit increase in the proportion of non-family households in a ZCTA is associated with fall of 0.24 percentage point in the homeownership rate of that ZCTA. A ZCTA, irrespective of its regional location, which has a higher proportion of family households (married couples etc.) will on average have a higher homeownership rate. From Table 6, *propNonFamily* has a standard deviation of 13.25 in my sample which indicates that a 1 standard deviation increase in this variable is expected to result in a decrease in homeownership rate, equal to 3.23 percentage points.

While both these demographic variables have a conditionally negative impact on the homeownership rate, *bachelors* has a net positive effect. *bachelors* represents the proportion of householders with at least a Bachelors' degree. Results from Model 4 in Table 5 support the prediction from hypothesis H1c. All else equal, one percentage point increase in the proportion of householders with a Bachelors' degree or higher in a ZCTA is associated with an increase of 0.24 percentage points in the homeownership rate of the ZCTA. This regression coefficient is also highly significant at the 0.001 level of significance. In standardized terms from Table 6, one standard deviation increase in *bachelors* leads to more than 5 percentage point increase in homeownership rate, conditional on all other included variables. Thus the data are consistent with the predictions in hypotheses H1a, H1b and H1c.

6.2 Analysis of Hypothesis 2

I use the proportion of housing structures with more than five units as a proxy for the density of the area. Existence of more multifamily structures indicates a higher proportion of renting and hence a lower homeownership rate. From Model 4 in Table 5, the coefficient on *moreThan5Units* is negative and statistically significant at the 0.001 level of significance. Conditional on all other included variables, a one unit increase in the proportion of multifamily structures in a ZCTA is associated with a

decrease of approximately half a percentage point in the homeownership rate of the ZCTA.

From Table 6, this is equivalent to more than 7 percentage points decrease in the homeownership rate for every one standard deviation increase in the proportion of multifamily structures. Thus, a ZCTA located in a central city is expected to have a lower homeownership rate compared to a similar ZCTA located in a suburban area. This confirms the prediction in hypothesis H2.

6.3 Analysis of Hypotheses 3

6.3.1 Price-rent Ratio and Homeownership Rate

Table 3 reports results of iterations to a basic model specification without including any regional terms. The results show that regression coefficient for *priceRentRatio* is negative and statistically significant at 0.001 level of significance. This represents the marginal impact of *priceRentRatio* of a ZCTA on the homeownership rate, conditional on all other variables. In Model 7, this is equal to -0.0672, i.e., for every one unit increase in the price-rent ratio across ZCTAs, the homeownership rate is expected to decrease by 0.07 percentage points. This is consistent with the prediction of H3a in accordance with the present-value approach to housing market.

This relationship changes when we consider the different nature of housing markets across regions. From Model 4 in Table 5, regression coefficient on *priceRentRatio* represents the effect of price-rent ratio on homeownership for the West. This is equal to -0.0018 and is not statistically different from zero. Since West is the reference category, regression coefficients on variables x_3d_1, x_3d_2 and x_3d_3 indicate the marginal impacts of price-rent ratio on homeownership rate for ZCTAs in Midwest, Northeast and South, respectively.

The coefficient on *priceRentRatio:mw* is equal to 0.2517 and is statistically significant at 0.001 level of significance. Consistent with hypothesis H3a, this means that assuming everything else is same across two ZCTAs including the regional location, the ZCTA with a higher price-rent ratio is expected to have a higher homeownership rate. This also implies that conditional on all other included variables, price-rent

ratio has a larger impact on the homeownership rate of a ZCTA in the Midwest relative to the West.

The Northeast also shows a similar result. Regression coefficient for *priceRentRatio:ne* is 0.0787 and is also statistically significant at 0.001 level of significance. The positive sign confirms the prediction of hypothesis H3a for the Northeast too. However, the differential impact of regional location on the relationship between price-rent ratio and homeownership rate in a ZCTA is statistically different and stronger in the Midwest than in the Northeast.

The regression coefficient on *priceRentRatio:so* is -0.0327. This is the marginal impact of price-rent ratio on the homeownership rate of a ZCTA in the South relative to one in the West. The net regression coefficient is -0.0345 which is not statistically significant. This is also not consistent with the prediction in hypothesis H3a. The net effects of price-rent ratio on homeownership rate for different regions can be seen from column 1 in Table 6. Conditional on everything else, if price-rent ratio increases by 1 unit in a ZCTA, the homeownership rate is expected to increase by 0.25 percentage points if the ZCTA is in the Midwest and by 0.08 percentage points if the ZCTA is in the Northeast. In standardized terms, this means that if price-rent ratio increases by one standard deviation in a ZCTA, the homeownership rate is expected to increase by 1.82 percentage points if the ZCTA is in the Midwest and by 0.87 percentage points if the ZCTA is in the Northeast.

6.3.2 Price-income Ratio and Homeownership Rate

From Model 4 in Table 5, regression coefficient on *priceIncRatio* represents the effect of price-income ratio on homeownership for the West. This is equal to -1.0158 and is statistically significant at 0.001 level of significance. Since West is the reference category, regression coefficients on variables x_4d_1, x_4d_2 and x_4d_3 indicate the marginal impacts of price-income ratio on homeownership rate for ZCTAs in Midwest, Northeast and South, respectively.

The coefficient on *priceIncRatio:mw* is equal to -0.7706 and is statistically significant at 0.001 level of significance. Consistent with hypothesis H3b, this means that

assuming everything else is same across two ZCTAs including the regional location, the ZCTA with a higher price-income ratio is expected to have a lower homeownership rate. This also implies that conditional on all other included variables, price-income ratio has a larger impact in absolute terms on the homeownership rate of a ZCTA in the Midwest relative to the West.

The Northeast also shows a similar result. Regression coefficient for *priceIncRatio:ne* is -0.8637 and is also statistically significant at 0.001 level of significance. The negative sign confirms the prediction of hypothesis H3b for the Northeast too. The differential impact of regional location on the relationship between price-income ratio and homeownership rate in a ZCTA is statistically different and stronger in absolute terms in the Northeast than in the Midwest. The regression coefficient on *priceIncRatio:so* (Model 4 in Table 5) is positive and statistically significant at 0.001 level of significance. This is the marginal impact of price-income ratio on the homeownership rate of a ZCTA in the South relative to one in the West.

The net effects of price-rent ratio on homeownership rate for different regions can be seen from column 1 in Table 6. Conditional on everything else, if price-income ratio increases by 1 unit in a ZCTA, the homeownership rate is expected to decrease by 1.02 percentage points if the ZCTA is in the West, by 1.79 percentage points if the ZCTA is in the Midwest and by 1.88 percentage points if the ZCTA is located in the Northeast. On the other hand, for a ZCTA located in the South, 1 unit increase in price-income ratio is associated with an increase of 0.06 percentage points in the homeownership rate of that ZCTA.

6.3.3 Regional Variation

The regional variation in homeownership rates is accounted for by the insertion of regional dummies into the regression equation. This is an estimate of the difference in homeownership rates, conditional on covariates included in the model i.e. how much of the regional variation is due to unobservables, and ideally this should be zero. In Table 7, I utilize this reasoning to measure the extent of collective variation that is explained by the included covariates. Model A is a baseline model with the

Table 7: **Measurement of Regional Variation**

Variable	Model A	Model B
intercept	67.9606***	86.2161***
hhUnder35yrs	—	-0.5086***
priceRentRatio	—	-0.0319***
priceIncRatio	—	-0.2464***
propNonFamily	—	-0.2529***
moreThan5Units	—	-0.5724***
bachelors	—	0.2515***
mw	10.2171***	7.7914***
ne	7.3113***	4.2088***
so	5.4165***	4.0782***
adj-R^2	0.0410	0.6637
SSD	222.26	121.69

Note 1: *, **, and *** indicate significance at the 0.1, 0.01, and 0.001 levels respectively.

Note 2: Sum of Squared Differences (SSD) is defined as the sum of squares of pairwise differences between regression coefficients of the dummy variables,

$$SSD = \sum_{i=0}^{2} \sum_{j=i+1}^{3} (d_i - d_j)^2$$

where d_0, d_1, d_2 and d_3 indicate West, Midwest, Northeast and South respectively. $d_0 = 0$ and d_1, d_2, d_3 are the regression coefficients for mw, ne and so respectively.

Note 3: For four regions, there are six distinct absolute pairwise differences which are $|d_0 - d_1|$, $|d_0 - d_2|$, $|d_0 - d_3|$, $|d_1 - d_2|$, $|d_1 - d_3|$ and $|d_2 - d_3|$.

homeownership rates regressed only on the regional dummy variables and *West* is the omitted category. Model B includes all the covariates without any interaction terms so as to focus only on the explained regional variation.

Comparing the magnitudes of dummy coefficients in the two models form Table 7, for instance, the difference between West and Midwest reduces by over 3 percentage points, and the difference between West and South falls by 1.3 percentage points. In another case, the disparity between the size of Northeast and South coefficients reduces from close to 2 percentage points to just about 0.13 percentage points.

To quantify the variability into a single measure, I define the Sum of Squared Differences (SSD) as the sum of squares of pairwise differences between regression coefficients on the dummy variables. The measure *SSD* reduces from 222.26 in Model A to 121.69 in Model B. If the coefficients on a regional dummy measures the differences among regions as a residual, it can be concluded that the set of included explanatory variables in my specification can explain approximately 44.2% variation in the homeownership rates across regions.

However, the residual component also captures the influence of important omitted variables or factors that are internal to the market process and harder to measure, but that are correlated with homeownership such as school district, crime rate and cultural disposition towards homeownership. As Eilbott and Binkowski (1985) point out, the decision to own rather than rent may be influenced by noneconomic factors. Insofar as taste differences, for example, reflect regional differences within a country, a desire for homeownership may be associated with living in, or growing up, in a particular are of the country. Several studies have identified specific regional differences in housing tastes. For example, Krumm (1987) finds that during the period of his study 1976-1979, the propensity toward homeownership is greatest in the rural West and Central regions, and in the urban and rural South, and lowest for locations throughout the Northeast.

7 Conclusion

This paper investigates the factors that explain cross sectional variation in homeownership rate across zip codes and across regions. Specifically, do the factors that explain intra-regional variation in homeownership rate have similar effects across regions? I also investigate, what is the differential impact of housing market variables related to home value, rent and household income across different regions?

The key contributions of this paper are threefold. Firstly, existing studies are based on a data from individual households or on aggregate data for only Metropolitan Statistical Areas (MSA). No investigation of the ownership outcomes has used aggregate data for a large number of zip codes. Secondly, the issue of regional disparity has largely been discussed using data from the cities or the metro areas and the rural areas have been left out of such studies.

My study accounts for the heterogeneity across submarkets within metro areas, micro areas and rural areas. I model the regional effect on homeownership rate through two channels, the unobserved locational effect and the effect through interaction with observed market variables. I use a sample of 28,851 zip codes from the aggregate 5-year data from 2011 ACS. I construct six key demographic and economic variables for each zip code from the ACS data and assign the regional membership to each of these.

Conditional on other included variables, a one percentage point increase in the proportion of relatively younger households and non-family households is associated with the same decrease in homeownership rate across all four regions. Similarly, a percentage point increase in the proportion of householders holding a Bachelors' degree or higher in a zip code is associated with an increase of 0.24 percentage points in the homeownership rate of that zip code.

Demographic variables influence the consumption demand for housing which is expected to be driven in the same way irrespective of the location. Nonetheless, a proxy for local density is included in the model to account for specific urban locations more heavily concentrated with multifamily housing structures. All else equal, one percentage point increase in the proportion of multifamily housing structures is

seen to be associated with a decline in homeownership by half a percentage point. Additionally, there are key housing market economic variables that influence the homeownership rate in an area.

I show that the four regions can be considered as four separate housing markets. I rely on the use of interaction terms between the market variables and regional dummy variables to establish this result. High statistical significance confirms my hypotheses. The regression coefficient on price-rent ratio for the Midwest is more than three times that for Northeast. This indicates that one percentage point increase in price-rent ratio in a zip code leads to only 0.08 percentage point increase in homeownership if the zip code is in the Northeast but to 0.25 percentage points increase if the zip code is in the Midwest.

Similarly, the price-income ratio, which is an indicator of relative affordability in a zip code has very different impact on homeownership depending on the regional location. South is the only region where price-income ratio is associated positively with homeownership rate, which is inconsistent with my hypothesis. However, the small magnitude (0.06 percentage points) may be indicative of homeownership becoming agnostic to home prices or income due to easier access to credit (Rosen et al. (2017)).

Results indicate that factors that are commonly associated with housing demand like household demographic profile and density are associated similarly with the aggregate homeownership rate in a zip code irrespective of the regional location of the zip code. However, variables related to housing prices have different effects on the homeownership rate within and across regions. While the specific empirical results pertain only to the unique time period of 2007-2011 and the sample data at the zip code level, the presence of these differences merit a more general discussion.

Region descriptive statistics (Table 1) report that both the coastal regions, West and Northeast, have the lowest homeownership rates, highest median household incomes and highest median home values. Much of this can be attributed to the various forms of supply side constraints in these regions. High-slope and coastal areas are severely land-constrained by their topography, rendering housing supply relatively inelastic. Furthermore, the extent of topographical constraints correlates

positively and strongly with regulatory barriers to development (Saiz (2008), Saiz (2010)). Gyourko et al. (2008) notes that the coastal states have the most highly regulated communities on average.

Land use regulations in the United States are widespread, largely under local control, and may be a major factor accounting for why housing appears to be in inelastic supply in many of the larger coastal markets. Zoning ordinances enacted by local governments greatly affect the costs, size and quantity of new single family homes built. In areas with stricter regulations, building costs typically increase and fewer homes are built, especially when compared with areas with fewer regulations. As land use regulations stifle new supply, land prices and construction costs rise, greatly reducing affordability and limiting homeownership. These supply side constraints are likely to manifest in the observed differential relationship between homeownership and housing prices variables.

Lastly, I interpret the regression coefficients on regional dummy variables to indicate any inherent locational effects not captured in my regression specification. The relative magnitudes of these dummy coefficients reduce significantly by including my selected covariates. When I combine the pairwise differences into a single metric, I find a decrease of 45.2% in this unobserved component. The existing level differences in regional dummies may be due to other unobserved factors that are likely to generate heterogenous inter-regional association with homeownership. For instance, average mortgage rates, household's marginal tax bracket and occupation profile may vary depending on the location. Spatial patterns of subprime lending reveal a disproportionately large market share in low income, low education and minority neighborhoods (Hershaff et al. (2005)). Similarly, Rosen et al. (2017) report that the declines in household income observed during the Great Recession correlate to the type of jobs lost. Mid-wage jobs paying between $14 and $21 per hour consisted of nearly 60% of the jobs lost during the recession. Composition of housing stock, cost of maintenance and depreciation and cultural preferences towards homeownership are some of the other unobserved factors that may either confound with observed determinants or have independent effects.

Therefore, my study creates an opportunity to investigate the changes that have

occurred in the housing market and consequently in the homeownership behavior around the Great Recession. Shiller (2007) reports that the increase in homeownership since 1994 was largely due to the remarkable housing boom. The period in my study belongs to the time immediately following the burst of housing bubble. Performing the same exercise for the time period 2012-2016 might generate insights into the changing attributes toward homeownership as well as structural changes in regional and national housing market.

References

Aaronson, D. (2000). A note on the benefits of homeownership. *Journal of Urban*.

Acolin, A., J. Bricker, P. Calem, and S. Wachter (2016). Borrowing constraints and homeownership. *American Economic Review 106*(5), 625–29.

Adelino, M., A. Schoar, and F. Severino (2018). Perception of house price risk and homeownership. Technical report, National Bureau of Economic Research.

Amior, M. and J. Halket (2014). Do households use home-ownership to insure themselves? evidence across us cities. *Quantitative Economics 5*(3), 631–674.

Baffoe-Bonnie, J. (1998). The dynamic impact of macroeconomic aggregates on housing prices and stock of houses: a national and regional analysis. *The Journal of Real Estate Finance and Economics 17*(2), 179–197.

Begley, J., L. Loewenstein, and P. S. Willen (2019). The price-rent ratio during the boom and bust: Measurement and implications.

Bernheim, B. D., D. M. Garrett, and D. M. Maki (2001). Education and saving:: The long-term effects of high school financial curriculum mandates. *Journal of public Economics 80*(3), 435–465.

Birkenmaier, J. and S. W. Tyuse (2006). Does homeownership education and counseling (hec) help credit scores? *Journal of social service research 32*(2), 81–103.

Bourassa, S. C. (1995). A model of housing tenure choice in australia. *Journal of Urban Economics 37*(2), 161–175.

Capozza, D. R. and P. J. Seguin (1996). Expectations, efficiency, and euphoria in the housing market. *Regional Science and Urban Economics 26*(3-4), 369–386.

Collins, J. M. (2009). Education levels and mortgage application outcomes: Evidence of financial literacy. discussion paper no. 1369-09. *Institute for Research on Poverty*.

Colton, K. W., K. Collignon, and U. Planning (2001). *Multifamily rental housing in the 21st century*. Joint Center for Housing Studies, Graduate School of Design [and] John F ....

Coulson, N. E. (1999). Why are hispanic-and asian-american homeownership rates so low?: Immigration and other factors. *Journal of Urban Economics 45*(2), 209–227.

Coulson, N. E. (2002). Regional and state variation in homeownership rates; or if california's home prices were as low as pennsylvania's would its ownership rate be as high? *The Journal of Real Estate Finance and Economics 24*(3), 261–276.

Cronin, D. and K. McQuinn (2016). Credit availability, macroprudential regulations and the house price-to-rent ratio. *Journal of Policy Modeling 38*(5), 971–984.

Crosby, L. A. and J. R. Taylor (1981). Effects of consumer information and education on cognition and choice. *Journal of Consumer Research 8*(1), 43–56.

Dıaz, A. and M. J. Luengo-Prado (2010). Homeownership: economic benefits. *The international encyclopedia of housing and home*, 439–445.

Dietz, R. D. and D. R. Haurin (2003). The social and private micro-level consequences of homeownership. *Journal of urban Economics 54*(3), 401–450.

DiPasquale, D. and E. Glaeser (1999). Incentives and social capital: Are homeowners better citizens?

Earley, F. (2004). What explains the differences in homeownership rates in europe? *Housing Finance International 19*(1), 25.

Eilbott, P. and E. S. Binkowski (1985). The determinants of smsa homeownership rates. *Journal of Urban Economics 17*(3), 293–304.

Elsinga, M. and J. Hoekstra (2005). Homeownership and housing satisfaction. *Journal of Housing and the Built Environment 20*(4), 401–424.

Fisher, L. M. and A. J. Jaffe (2003). Determinants of international home ownership rates. *Housing finance international 18*(1), 34–37.

Gallin, J. (2008). The long-run relationship between house prices and rents. *Real Estate Economics 36*(4), 635–658.

Gelain, P. and K. J. Lansing (2014). House prices, expectations, and time-varying fundamentals. *Journal of Empirical Finance 29*, 3–25.

Gillingham, R. and R. Hagemann (1983). Cross-sectional estimation of a simultaneous model of tenure choice and housing services demand. *Journal of Urban Economics 14*(1), 16–39.

Glaeser, E. L. (2007). The economics approach to cities. Technical report, National Bureau of Economic Research.

Glaeser, E. L. and J. Gyourko (2002). The impact of zoning on housing affordability. Technical report, National Bureau of Economic Research.

Glaeser, E. L. and J. M. Shapiro (2003). The benefits of the home mortgage interest deduction. *Tax policy and the economy 17*, 37–82.

Goodman, A. C. (1988). An econometric model of housing price, permanent income, tenure choice, and housing demand. *Journal of Urban Economics 23*(3), 327–353.

Goodman, L. S. and C. Mayer (2018). Homeownership and the american dream. *Journal of Economic Perspectives 32*(1), 31–58.

Goswami, G., S. Tan, and M. Waisman (2014). Understanding the cross-section of the us housing bubble: The roles of lending, transaction costs, and rent growth. *Journal of Financial Stability 15*, 76–90.

Grant, C. (2007). Estimating credit constraints among us households. *Oxford Economic Papers 59*(4), 583–605.

Green, R. K., M. J. White, et al. (1997). Measuring the benefits of homeowning: Effects on children. *Journal of urban economics 41*(3), 441–461.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Güriş, S., E. Çağlayan, and T. Ün (2011). Estimating of probability of homeownership in rural and urban areas: Logit, probit and gompit model. *European Journal of Social Sciences 21*(3), 405–411.

Gwin, C. R. and S.-E. Ong (2008). Do we really understand homeownership rates? an international study. *International Journal of Housing Markets and Analysis*.

Gyourko, J. and P. Linneman (1996). Analysis of the changing influences on traditional households' ownership patterns. *Journal of Urban Economics 39*(3), 318–341.

Gyourko, J. and A. Saiz (2006). Construction costs and the supply of housing structure. *Journal of regional Science 46*(4), 661–680.

Gyourko, J., A. Saiz, and A. Summers (2008). A new measure of the local regulatory environment for housing markets: The wharton residential land use regulatory index. *Urban Studies 45*(3), 693–729.

Halket, J., L. Nesheim, and F. Oswald (2015). The housing stock, housing prices, and user costs: the roles of location, structure and unobserved quality. Technical report, cemmap working paper.

Hattapoglu, M. and I. Hoxha (2014). The dependency of rent-to-price ratio on appreciation expectations: an empirical approach. *The Journal of Real Estate Finance and Economics 49*(2), 185–204.

Haurin, D. R., P. H. Hendershott, and S. M. Wachter (1996). Borrowing constraints and the tenure choice of young households. Technical report, National bureau of economic research.

Haurin, D. R., S. M. Wachter, and P. H. Hendershott (1995). Wealth accumulation and housing choices of young households: An exploratory investigation. Technical report, National Bureau of Economic Research.

Henderson, J. V. and Y. M. Ioannides (1983). A model of housing tenure choice. *The American Economic Review 73*(1), 98–113.

Hershaff, J., S. Wachter, K. Russo, and S. M. Wachter (2005). Subprime lending: neighborhood patterns over time. In *Proceedings from Promises and Pitfalls Conference, Federal Reserve System Community Affairs Research Conference*.

Hilber, C. A. and Y. Liu (2008). Explaining the black–white homeownership gap: the role of own wealth, parental externalities and locational preferences. *Journal of Housing Economics 17*(2), 152–174.

Hilgert, M. A., J. M. Hogarth, and S. G. Beverly (2003). Household financial management: The connection between knowledge and behavior. *Fed. Res. Bull. 89*, 309.

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

Holmes, A. and P. Horvitz (1994). Mortgage redlining: Race, risk, and demand. *The Journal of Finance 49*(1), 81–99.

Hullgren, M. and I.-L. Söderberg (2013). The relationship between consumer characteristics and mortgage preferences. *International Journal of Housing Markets and Analysis*.

Ioannides, Y. M. and S. S. Rosenthal (1994). Estimating the consumption and investment demands for housing and their effect on housing tenure status. *Review of Economics and Statistics 76*(1), 127–141.

Ionescu, F. and N. Simpson (2010). Credit scores and college investment.

Jurgilas, M. and K. J. Lansing (2012). Housing bubbles and homeownership returns. *FRBSF Economic Letter 19*.

King, M. A. (1980). An econometric model of tenure choice and demand for housing as a joint decision. *Journal of public Economics 14*(2), 137–159.

Kishor, N. K. and J. Morley (2015). What factors drive the price–rent ratio for the housing market? a modified present-value analysis. *Journal of Economic Dynamics and Control 58*, 235–249.

Krumm, R. (1987). Intertemporal tenure choices. *Journal of Urban Economics 22*(3), 263–275.

Lamont, O. and J. C. Stein (1997). Leverage and house-price dynamics in us cities. Technical report, National Bureau of Economic Research.

Leamer, E. (2002). Bubble trouble? your home has a p/e ratio too. *UCLA Anderson Forecast*, 1–11.

Lee, S. W. and D. Myers (2003). Local housing-market effects on tenure choice. *Journal of Housing and the Built Environment 18*(2), 129–157.

Lerbs, O. W. and C. A. Oberst (2014). Explaining the spatial variation in homeownership rates: Results for german regions. *Regional Studies 48*(5), 844–865.

Linneman, P., I. F. Megbolugbe, S. M. Wachter, and M. Cho (1997). Do borrowing constraints change us homeownership rates? *Journal of Housing Economics 6*(4), 318–333.

Lusardi, A. (2003). Planning and saving for retirement. Technical report, Working paper. Dartmouth College.

Lusardi, A. (2008). Household saving behavior: The role of financial literacy, information, and financial education programs. Technical report, National Bureau of Economic Research.

Lusardi, A. and O. S. Mitchelli (2007). Financial literacy and retirement preparedness: Evidence and implications for financial education. *Business economics 42*(1), 35–44.

Mayo, S. K. (1981). Theory and estimation in the economics of housing demand. *Journal of Urban Economics 10*(1), 95–116.

Myers, D., G. Painter, and J. Zissimopoulos (2016). Education and the intergenerational transmission of homeownership. Technical report, Working paper). Washington, DC: Fannie Mae. Retrieved from http://www ....

Nissan, E. and J. E. Payne (2013). A simple test of σ-convergence in us housing prices across bea regions. *Journal of Regional Analysis and Policy 43*(1100-2016-90116), 178–185.

Ortalo-Magné, F. and A. Prat (2014). On the political economy of urban growth: Homeownership versus affordability. *American Economic Journal: Microeconomics 6*(1), 154–81.

Ortalo-Magné, F. and S. Rady (2002). Tenure choice and the riskiness of non-housing consumption. *Journal of Housing Economics 11*(3), 266–279.

Painter, G., S. Gabriel, and D. Myers (2001). Race, immigrant status, and housing tenure choice. *Journal of Urban Economics 49*(1), 150–167.

Painter, G. and C. L. Redfearn (2002). The role of interest rates in influencing long-run homeownership rates. *The Journal of Real Estate Finance and Economics 25*(2-3), 243–267.

Peng, C.-W. and T. T. Yang (2015). The impact of price-to rent ratios on homeownership rates.

Phang, S. Y. (2009). Affordable homeownership policy: implications for housing markets.

Pickvance, C. G. (1974). Life cycle, housing tenure and residential mobility: A path analytic approach. *Urban Studies 11*(2), 171–188.

Priceonomics (2017, February 16). Which college degrees produce the most (and least) financially responsible students? https://priceonomics.com/which-college-degrees-produce-the-most-and-least. (Accessed on 07/19/2020).

Proxenos, S. (2002). Homeownership rates: a global perspective. *Housing finance international 17*(2), 3–7.

Quercia, R. G., G. W. McCarthy, and S. M. Wachter (2003). The impacts of affordable lending efforts on homeownership rates. *Journal of Housing Economics 12*(1), 29–59.

Reichert, A. K. (1990). The impact of interest rates, income, and employment upon regional housing prices. *The Journal of Real Estate Finance and Economics 3*(4), 373–391.

Robinson, M., G. M. Scobie, and B. Hallinan (2006). Affordability of housing: concepts, measurement and evidence. Technical report, New Zealand Treasury Working Paper.

Rosen, K. T., D. Bank, H. Belfor, A. Eckstein, A. Gade, and M. Tcheau (2017). Hurdles to homeownership: Understanding the barriers.

Rosen, K. T., D. Bank, A. Eckstein, M. Stern, and M. Tcheau (2017). Homeownership in crisis: Where are we now?

Rossi, P. H. and E. Weber (1996). The social benefits of homeownership: Empirical evidence from national surveys. *Housing policy debate 7*(1), 1–35.

Saiz, A. (2008). On local housing supply elasticity. *Available at SSRN 1193422*.

Saiz, A. (2010). The geographic determinants of housing supply. *The Quarterly Journal of Economics 125*(3), 1253–1296.

Shiller, R. J. (2007). Understanding recent trends in house prices and home ownership. Technical report, National Bureau of Economic Research.

Sinai, T. and N. S. Souleles (2005). Owner-occupied housing as a hedge against rent risk. *The Quarterly Journal of Economics 120*(2), 763–789.

Struyk, R. J. and S. Marshall (1974). The determinants of household home ownership. *Urban Studies 11*(3), 289–299.

Turner, T. M. and H. Luea (2009). Homeownership, wealth accumulation and income status. *Journal of Housing Economics 18*(2), 104–114.

Zhou, Y. (2013). *The Changes in Young Households' Tenure Choice across Time: Evidence from American Housing Survey.* Ph. D. thesis, Tufts University.

A Data Definitions

Metropolitan Statistical Area (MSA). A census-designated area with at least one urbanized area of 50,000 or more inhabitants.

Micropolitan Statistical Area. A census-designated area with at least one urban cluster of at least 10,000 but less than 50,000 population.

Housing Unit. A housing unit is a house, an apartment, a group of rooms, or a single room occupied or intended for occupancy as separate living quarters. Separate living quarters are those in which the occupants do not live and eat with other persons in the structure and which have direct access from the outside of the building or through a common hall. Living quarters of the following types are excluded from the housing unit inventory: Dormitories, bunkhouses, and barracks; quarters in predominantly transient hotels, motels, and the like, except those occupied by persons who consider the hotel their usual place of residence; quarters in institutions, general hospitals, and military installations except those occupied by staff members or resident employees who have separate living arrangements.

Occupied Housing Units. A housing unit is occupied if a person or group of persons is living in it at the time of the interview or if the occupants are only temporarily absent, as for example, on vacation. The persons living in the unit must consider it their usual place of residence or have no usual place of residence elsewhere. The count of occupied housing units is the same as the count of households.

Householder. The householder refers to the person (or one of the persons) in whose name the housing unit is owned or rented or, if there is no such person, any adult member, excluding roomers, boarders, or paid employees. If the house is jointly owned by a married couple, either the husband or the wife may be listed first, thereby becoming the reference person, or householder, to whom the relationship of the other household members is recorded. One person in each household is designated as the "householder."

Number of Housing Units in Structure. A structure is a separate building which either has open space on all four sides or is separated from other structures by dividing walls that extend from ground to roof. The count of housing units in a structure is the total number of units in the structure, both occupied and vacant units. In the tabulations, occupied mobile homes or trailers, tents, and boats are included in the category one housing unit in structure.

Contract Rent. For renter-occupied units, the contract rent is the monthly rent agreed upon regardless of any furnishings, utilities, or services that may be included. The median monthly rent is the rent which divides the series into two equal parts, one-half of the units with rents higher than the median and the other half with rents lower than the median.

Value. Value is the respondent's estimate of how much the property would sell for on the current market. For vacant units, value is the sales price asked for the property at the time of the interview and may differ from the price at which the property is sold. The median value or sales price is the amount which divides the series into two equal parts, one-half of the units with values higher than the median and the other half with values lower than the median.

Homeownership Rate. The proportion of households that are owners is termed the homeownership rate. It is computed by dividing the number of households that are owners by the total number of households. The formula is as follows:

$$\textit{Homeownership Rate (\%)} = \left[\frac{\textit{Owner occupied housing units}}{\textit{Total occupied housing units}}\right] * 100$$

Family. A family is a group of two persons or more (one of whom is the householder) related by birth, marriage, or adoption and residing together; all of such persons (including related subfamily members) are considered as members of one family.

Non-family Householder. A non-family householder is a person maintaining a household while living alone or with non-relatives only.

B Census Regions

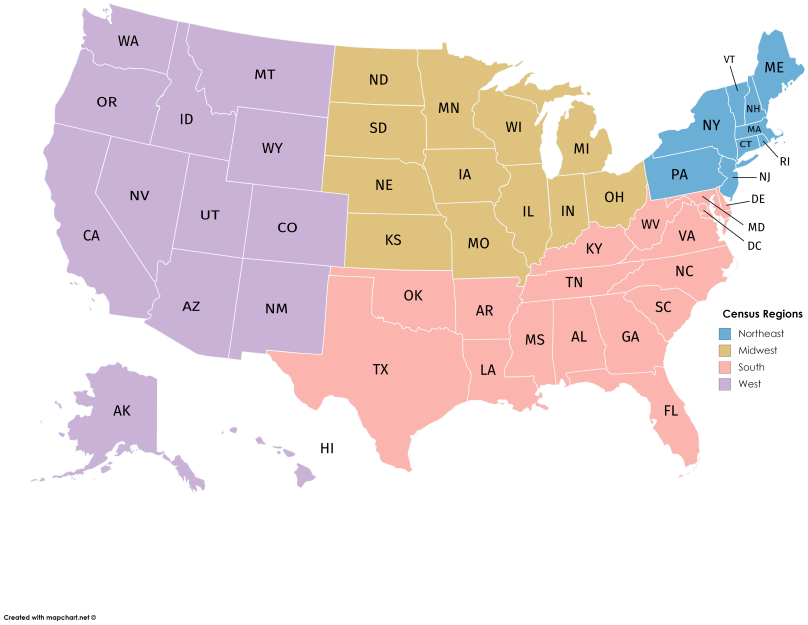


Figure 4: **Four Census Regions**

Source: https://mapchart.net/usa.html
Notes:
Northeast: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont, New Jersey, New York, Pennsylvania.
Midwest: Illinois, Indiana, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota.
South: Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma, Texas.
West: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming, Alaska, California, Hawaii, Oregon, Washington.

C Coefficient of Variation

Region	Median Home Value	Median Household Income	Value-Income Ratio
West	0.7120	0.4287	1.3028
Midwest	0.5721	0.3274	0.6437
Northeast	0.7382	0.4479	0.6602
South	0.7998	0.4613	1.5863

Source: 2007-11 American Community Survey, U.S. Census Bureau
Note: Values represent Coefficient of Variation (CV) for region i computed as,

$$CV_i = \frac{\sigma_i}{\bar{X}_i}$$

where $\bar{X}_i$ and σ_i are the sample mean and sample standard deviation of that variable for region i, respectively.

Chapter 2:
Changes in Homeownership Behavior around the Great Recession

Shikha Agarwal

Abstract

This paper investigates the impact of the housing bubble on homeownership during the bust and the initial phase of economic recovery on three dimensions. Is there a structural break in the national housing market in 2011-2012? Did regional disparities due to unobserved region-specific factors significantly decrease during the recovery period of 2012-2016? Finally, since most of the run-up in homeownership was due to relaxed credit standards, how did the response of homeownership to relative affordability change before and after the crash in house prices? I find that the housing markets were indeed very different before and after the housing crash at both the national and sub-national level. Not only did the homeownership response change for all included variables and for all regions, but also the recovery phase can better explain collective variation across regions and across zip codes. I also find that the homeownership response of households to relative affordability became stronger as a result of both personal experience of foreclosures as well as tighter lending standards.

1 Introduction

The rate of homeownership is traditionally interpreted as a major indicator of economic welfare. Together, housing services and residential fixed investment spending constitute the majority of housing related expenditure nationwide, and account for more than 15% of GDP. Beyond improving overall economic conditions, homeownership is beneficial to both individual households and society as a whole. Homeownership is associated with a high level of wealth accumulation and economic freedom (Turner and Luea (2009)). Homeowners also tend to be more satisfied with their housing situation than tenants (Elsinga and Hoekstra (2005)). Several studies even report a positive influence of homeownership on social cohesion and stability (Rossi and Weber (1996), DiPasquale and Glaeser (1999), Dietz and Haurin (2003)).

The promotion of homeownership has for decades been one of the most popular political objectives in the United States. The decade through 2004 witnessed robust economic expansion and a policy environment promoting broad access to homeownership (Rosen et al. (2017)). As Mattews and Robert (2016) note, the housing bubble of 2007 was a creation of both market forces and government policy mistakes. Subsequently, in the aftermath of the Great Recession, homeownership fell with tighter credit conditions, problematic student loan debt, stagnant real incomes, and perhaps a subtle change in attitudes toward homeownership.

I aim to investigate the impact of the housing bubble on homeownership during the bust and the initial phase of economic recovery. Specifically, is there a structural break in the housing market in 2011-2012, both nationwide and at the regional level? Painter and Redfearn (2002) emphasize that regional disparities within the United States underscore the role of local factors in influencing homeownership rate. Since part of the run-up in homeownership was due to relaxed credit standards (Goodman and Mayer (2018)), how has the homeownership response to relative affordability changed before and after the crash of house prices?

I use aggregate 5-year data from the 2011 and 2016 American Community Survey (ACS) at the zip code level. These contain 'period' estimates of data collected over the time period 2007-2011 and 2012-2016, respectively. My choice of time periods

is based on the developments of the housing bubble and the resulting credit crisis. Home prices reached their peak in the second quarter of 2006 and did not start recovering until the end of 2011. The first time period of 2007-2011 is the bust phase in the housing market and the second period of 2012-2016 can be thought of as the phase of initial recovery. My sample consists of a panel of 29,575 zip codes for each of the two time periods, consisting of both urban and rural areas. I construct six key demographic and economic variables for each zip code from the ACS data. Additionally, I assign the regional location to each zip code based on U.S. Census codes.

I follow a methodology similar to Agarwal (2020), who studies cross-sectional variation in homeownership rates and develops a metric to measure unexplained variation across regions. The region can affect the homeownership rate of an area not only through its unobserved locational effect but also through interaction with demographic and market variables. I use six explanatory variables, three dummy variables and relevant interaction terms to explain variation in homeownership rates across regions and over the time period, 2007-2016. I evaluate the econometric model for both time periods and test for joint equality of the coefficients using the Chow test (Chow (1960)). Next, I define a measure of the regional variation in homeownership explained by included explanatory variables and create a test-statistic to compare this across time. I utilize the basic nonparametric bootstrap algorithm to generate distribution of this custom test statistic and test for a significant difference using a paired t-test.

The results are consistent with the predictions of my hypotheses. I test for joint equality of regression coefficients from a linear regression model fit on the data for each time period. The Chow test rejects the null hypothesis of no structural break at a high level of statistical significance. This implies that the response of homeownership to the variables included in my model is indeed significantly different in the recovery period relative to the crisis period. The housing crisis not only altered lending and banking practices but also households' behavior and attitude towards homeownership. The idea of a return to more 'normal' times is also analysed by testing for regional differences in homeownership due to unobserved factors. I define

a measure of collective variation and the reduction in this measure across time to build a test statistic.

Using a bootstrapped distribution of this statistic, I find that my included factors have a more homogeneous effect on the homeownership rate across regions in the recovery period than in the crisis period. Next, I go a step further to focus only on the response of homeownership rate in a zip code to the ratio of median home value to median household income. Since there was a dramatic shift from lax lending standards to a much tighter credit regime, I expect homeownership to be more responsive in the recovery period. I find that an increase in this affordability measure leads to a much larger decrease in homeownership rate during the recovery phase compared to the response in the bust phase. This result is also statistically significant and consistent for all regions.

This paper improves upon the existing empirical studies in three main ways. Firstly, existing studies are based on data from individual households or on aggregate data for only Metropolitan Statistical Areas (MSA). No investigation of the ownership outcomes across regions and across time has used aggregate data for a large number of zip codes. This paper uses data for almost all zip codes covered by the 5-year American Community Survey (ACS). This enables me to account for heterogeneity across submarkets within MSAs, not just across MSAs.

Secondly, the issue of regional disparity has largely been discussed using data from the cities or the metro areas and the rural areas have been left out of such studies. In fact, Glaeser and Gyourko (2002) point out that smaller datasets which feature only large metropolitan areas may oversample highly regulated cities and underrepresent the bulk of American housing. Lastly, I create a custom metric and test statistic to confirm the change in the housing market at the regional level. This serves as a way to isolate regional differences from zip code level data and focus on only collective inter-regional disparity in homeownership rates that exists due to unobserved or excluded factors.

The remainder of the paper proceeds as follows. Section 2 reviews prior literature and develops the hypotheses of interest. Section 3 describes the data source and sample construction. Section 4 describes the econometric models and test statistics.

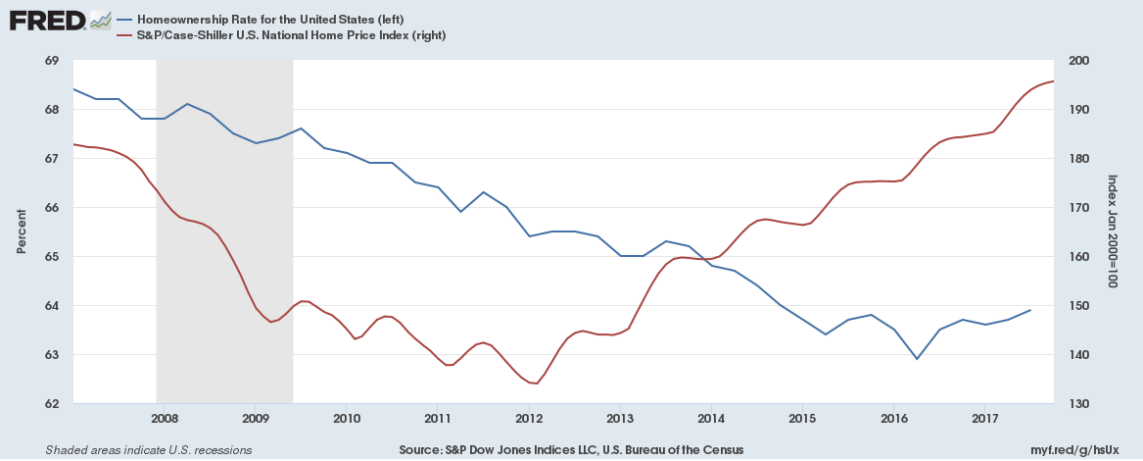


Figure 1: **Homeownership and Home Price Trend, 2007-2017**

Source: https://fred.stlouisfed.org/
Note: Figure shows the nationwide trend of quarterly homeownership rate(%) on the left y-axis and the Case-Shiller U.S. National Home Price Index (Index Jan 2000=100) on the right y-axis.

Section 5 presents the results and discusses robustness. Section 6 provides concluding remarks.

2 Literature & Hypotheses Development

2.1 Fundamental Changes in the Housing Market

Figure 1 depicts the trajectory of the Case-Shiller 20-City Price Index and the U.S. homeownership rate over the time period 2007-2016. During the period 2007-2011, both the Case-Shiller Index and the homeownership rate were falling on average across the nation. However, the relative movement of these two measures reversed direction in 2012. While home prices started recovering, homeownership remained low and on a downward trend. This indicates a stark change in the housing market around the cusp of 2011-12.

The homeownership rate increased during the housing boom through the mid-2000s, fueled by an extremely loose credit environment and sustained home price

appreciation (Rosen et al. (2017)). During the 1990s and early 2000s, the national homeownership rate skyrocketed to 69.2%, adding approximately 11.3 million new owner households nationwide from 1994 through 2004. This trend shifted rapidly, however, with the national homeownership rate plummeting over the following decade in the wake of the foreclosure crisis and great recession, throughout the early stages of recovery. As of 2016, the national homeownership rate reached an annual average of 63.4%, which marked the lowest level in more than 50 years and a decline of 5.6 percentage points compared with the pre-recession peak.

The National Bureau of Economic Research (NBER) declared the U.S. to have entered in a recession beginning December 2007 based on key economic indicators (WSJ (2008)). House prices fell about 9% in 2007 and stock market indices that peaked in October 2007 began to trend downward (Bricker et al. (2011)). O'Neill and Xiao (2012) reference the term "new normal" to describe the changed U.S. economic landscape. At the micro level, U.S. households were deleveraging their debt balances and increasing their rate of precautionary savings as a percentage of household income. In addition, more than half of working adults reported some measure of unemployment, such as a cut in pay, a reduction in work hours, or involuntary part-time employment (Taylor et al. (2010)). The financial crisis not only caused a deep recession and a sharp rise in unemployment, it also took a heavy toll on household financial positions.

During the Great Recession, the U.S. economy experienced its sharpest drop in household consumption expenditures in the postwar period (De Nardi et al. (2011); Petev et al. (2011)). Consumption declined precipitously across all categories, not just durables, from its peak in early 2007 to its trough roughly five years later. This rapid fall occurred at the end of a decade of markedly above-trend growth. Mian et al. (2013) associate these atypical aggregate consumption dynamics to the extraordinary swings in U.S. housing net worth that occurred since the end of the 1990s. O'Neill and Xiao (2012) find evidence that the financial crisis was associated with pre- and post-differences in financial practices. As Glumov (2013) reports, household net worth dropped, largely due to sharp declines in stock market valuations and in homeowner equity. He finds that investors became more risk-averse as the riskier assets in their

portfolios declined substantially in value, especially relative to other less volatile assets.

Ozdemir (2020) examines the link between housing market common factors, business and interest rate business cycles, and finds that the credit market factor switches to a low growth regime over the Great Recession. His results support the view that the incorporation of regime shifts in the housing market and credit market may contribute to improving our understanding of boom-bust cycles of the general economy. Cukierman (2013) shows that the global financial crisis resurrected concerns about financial stability, and underlined the importance of the central bank's lender-of-last-resort function and the use of unconventional monetary policy instruments. Barth et al. (2013) also show that the pre-crisis trend of loosening bank regulation reverted in the aftermath of the crisis with governments and regulators tightening their regulatory regimes.

Tan and Cheong (2016) reinforce the idea that the U.S. housing market underwent regime shifts between alternate stable states consistent with the observed signals. They find that the U.S. housing market experienced six different regimes during the period 1975-2011, last of which is a structural break in 2007Q1-2007Q2. The first period is attributed to a large historic rise in the subprime proportion of mortgages issued in the fourth quarter of 2003. 2007 marks the onset of the Great Recession, the second period. If their study was extended beyond 2011, I would expect that the recovery period after the recession would constitute another stable state marked by changes in relationship between homeownership and economic variables.

Nneji et al. (2013) discuss the relationship between the housing market and its macroeconomic determinants may be regime-varying, and so accounting for the housing market cycle is critical when examining its response to external macroeconomic factors. They find that the housing market becomes completely insensitive to changes in the economic variables in the crash regime, with prices generally being more sensitive during housing booms.

All these studies indicate that the Great Recession induced a fundamental break not only in mortgage lending practices but also in the household consumption, interest rate business cycles and monetary policy. As Brueckner et al. (2012) note, the

spectacular run-up and subsequent collapse of U.S. housing prices were watershed events in the real-estate history and no previous brief period witnessed such dramatic price escalation.

Based on this, I expect the following hypothesis to hold true,

H1a: The relationship between zip code level homeownership rate and its determinants changed in the recovery period (2012-2016) compared to the period of the housing crisis (2007-2011) due to changes in the household behavior and housing market dynamics. For instance, during the recession a growing number of residents doubled up on housing, moved back in with parents and delayed household formation, thus reducing the homeownership rate (Rosen et al. (2017)). This indicates that, for example, age of the householder may not be a strong determinant of homeownership during the recession but is potentially important during non-recessionary times.

All regions of the country witnessed different extent of decline in homeownership rate. Rosen et al. (2017) report that households in the Midwest and the Northeast regions, where the boom in construction was less pronounced and home price appreciation was generally more moderate through the early and mid-2000s, were the most resilient, leading to smaller declines in homeownership. Meanwhile, homeownership among households in the West and South regions proved most affected by the housing crisis. Among the four major Census regions, volatility in the homeownership rate was greatest in the West region, with the largest concentration of foreclosures. Northeast was the least volatile region of the country.

The rise in housing prices and the subsequent fall did not occur uniformly across the country (Baily et al. (2008), Bhattacharya and Kim (2011)). While the factors encouraging price increases applied broadly (especially the low interest rates), the impact on prices and the extent to which a bubble developed also depended largely on local conditions.

Aalbers (2009) investigates cross-sectional effects of the mortgage crisis and finds a great deal of variation not only in default and foreclosure rates, but also in regulation. He concludes that housing bubbles, faltering economies and regulation together have shaped the geography of the financial crisis at the state and city level. This indicates the possibility that homeownership responses might have changed at the sub-

national level i.e. can demographic and market variables explain more inter-regional variation after the crisis? Nissan and Payne (2013) examines whether convergence exists with respect to housing prices across the 50 states, especially in light of the housing market collapse and re-emergence of the housing market after the crisis. The results reveal a great deal of variation across states and regions. Miles (2015) shows that the extent of regional integration has important policy implications. In addition, understanding the interaction between different regional housing markets may improve forecasts of home values. However, Clark and Coggin (2009) finds that the evidence for regional convergence is mixed.

All four regions had varied experiences during the housing crash due to differences in regulation and underlying nature of housing market. After accounting for demographic and economic determinants of homeownership, the residual inter-regional differences exist due to difference in unobserved factors. It is reasonable to expect that in a recovering housing market, fundamental factors should be able to account for a higher proportion of inter-regional variation relative to the bust phase. Thus, I predict the following hypothesis to hold true,

H1b: Conditional on demographic and market variables, during the recovery in the housing market, we observe less regional variation due to unobserved factors.

2.2 Effect of Lending Practices on Homeownership Rate

The homeownership rate increased during the housing boom through the mid-2000s, fueled by an extremely loose credit environment and sustained home price appreciation. The Federal National Mortgage Association (FNMA), commonly called as Fannie Mae and the Federal Home Loan Mortgage Corporation (or Freddie Mac) are two Government Sponsored Enterprises (GSEs). They were created by Congress to provide liquidity, stability and affordability to the mortgage market and support the nation's housing finance system[1].

As house prices peaked in January 2006 and then rapidly declined, over a third

[1]https://www.fhfa.gov/SupervisionRegulation/FannieMaeandFreddieMac/Pages/About-Fannie-Mae—Freddie-Mac.aspx

of U.S. homes with mortgages fell "underwater" (Acolin et al. (2016)). In response to high foreclosure rates, mortgage underwriting was tightened over the period 2008-2013 beyond historic norms (Goodman et al. (2015)). Goodman (2018) states that mortgage credit is defined by a number of factors, including credit scores, loan-to-value (LTV) ratios, debt-to-income (DTI) ratios, and the type of mortgage loan the borrower takes out. He combines these factors into a single measure of market risk. He finds that the market risk at the beginning of 2017 is less than one-third of the risk in 2006-2007, a period when credit standards were too loose. Gwartney et al. (2008) identified relaxed mortgage lending standards as one of the key factors leading to the housing bubble and credit crisis.

Debt-to-income (DTI) ratio is one of a number of risk factors a lender considers in assessing whether to offer a mortgage to a borrower. Federal Housing Finance Agency (FHFA) defines DTI as the percentage of a borrower's income that is devoted to debt. Generally, it is calculated by dividing the total monthly debt expense by the total monthly income at the time a mortgage is originated or modified (WPR-2019-002). Fannie Mae and Freddie Mac consider a borrower's DTI, along with other factors, to determine if a mortgage is eligible for purchase. Fannie Mae explains that, as the DTI increases, "the level of risk also tends to increase; and a high ratio," along with other high-risk factors, creates higher risks associated with the borrower. Freddie Mac states that a borrower with a higher DTI "increases the probability a borrower may be unable to meet all their obligations at some point in the future."

Prior to 2008, both GSEs purchased mortgages with DTI ratios up to 65%. The housing market experienced a significant downturn in 2008, and the GSEs adjusted various underwriting requirements, including DTI. In 2008 through early 2010, Fannie Mae and Freddie Mac each implemented a series of adjustments to DTI limits. In May 2008, Fannie Mae introduced additional requirements for mortgages with DTI ratios between 55%-65%. Soon after, in January 2009, Freddie Mac reduced the maximum DTI to 55% and by early 2010, both GSEs reduced their DTI limit to 50%. In 2014, the Consumer Finance Protection Bureau established a 43% debt-to-income requirement (Rosen et al. (2017)). This high income requirement has been a significant barrier to securing a mortgage in the initial years of housing recovery.

Down payment or loan-to-value (LTV) requirements also changed substantially in the years following the Great Recession. Historically, most mortgages were 30-year fixed rate loans requiring a down payment of at least 20% of the property value and mortgage insurance if the 20% down payment requirement was not met (Holt (2009)). The borrowers also had to prove that their income was sufficient to ensure that the monthly mortgage payments would be manageable. Monthly payments generally were required to be no more than 36% of monthly income (underwriting ratio).

Baily et al. (2008) notes that a deterioration in lending standards can be dated back to 2004 or 2005. Families that lacked the income and down payment to buy a house under the specific mortgage terms of Fannie Mae and Freddie Mac were encouraged to take out a mortgage that had a very high loan to value ratio, perhaps as high as 100% meaning that they started with no initial equity and thus no true financial stake in the house. By 2006 many mortgages had no down payment and no verification of income (Mattews and Robert (2016)). Such borrowing typically requires a high interest rate and high monthly payments, that likely violate the usual rules on the proportion of household income needed to service the debt. As default rates began to soar in 2006 and 2007, banks began to tighten their standards and to require larger down payments. As of the second quarter of 2016, the average down payment for conventional 30-year fixed rate purchase mortgages was more than 17% (Rosen et al. (2017)).

In addition to the change in DTI and LTV requirements, other mortgage financing practices also changed after the Recession. A major cause of the foreclosure crisis was the widespread use of risky alternative mortgage products such as Adjustable Rate Mortgage (ARM) loans, interest only loans, negative amortization mortgages, balloon loans and hybrid ARMs. Pavlov and Wachter (2009) show that the supply of aggressive lending instruments temporarily increased the asset price in the underlying market by making it more attractive for agents to own or by relaxing their borrowing constraint, or both. Schloemer (2006) finds that compared with fixed-rate mortgages, ARMs are more than twice as likely to default.

As per FHFA (2019), a "subprime" borrower refers to mortgage applicants considered to be the least creditworthy because of low credit scores or uncertain income

prospects. Alt-A borrowers, deemed a bit less risky but not quite prime, had better credit scores but little to no documentation of income. Historically, mortgages taken out by the lower-income households often did not conform to Fannie and Freddie's strict standards for down payment requirements and income requirements (Holt (2009)). Baily et al. (2008) describe the shifting composition of mortgage lending and the erosion of lending standards. In 2004-06, the share of subprime and home equity lending expanded greatly. Prime mortgages dropped to 64% of the total in 2004 from 85% in 2003, 56% in 2005 and 52% in 2006, meaning that nearly half of mortgage originations in 2006 were subprime, Alt-A or home equity.

Ashcraft and Schuermann (2008) summarise that the boom in mortgage borrowing was sustained by low interest rates, easier lending practices and lack of due diligence. A substantial portion of the former homeowners who are now perceived as particularly risky after losing a home to foreclosure, may never have gone through a foreclosure if they had taken out a conventional 15 or 30-year fixed rate loan, as opposed to a riskier alternative mortgage product such as interest only mortgages, negative amortizing mortgages and ARMs. During the peak of the housing boom in 2005, ARMs accounted for 42% of new mortgage origination, according to the Urban Institute (2016). This number fell sharply following the crisis, decreasing to the low-1% range as of 2016.

U.S. housing and mortgage markets became stressed during 2007 and 2008 as a result of significant house price declines and the weakening economy. Since the financial crisis, the single-family mortgage credit market tightened dramatically, compounding the effects of lower wage growth and unemployment, and leading to a lower than average pace of new entrants into the homeownership market (Rosen et al. (2017)). Following the end of the Great Recession, new rules and standards on lenders contributed to the constrained credit environment and substantially inhibited lending even to the most creditworthy households. Acolin et al. (2016) find that tightened borrowing constraints have a substantial negative impact on the probability of becoming a homeowner in the aftermath of the Great Recession.

Based on this, I investigate the following hypothesis,

H2: Conditional on other factors, relationship between homeownership rate and the

ratio of median home value to median household income in a zip code became stronger during the recovery phase of 2012-2016 compared to the bust phase of 2007-2011 for all regions.

Moreover, the marginal effects for zip codes in the Midwest, Northeast and South also increased relative to the West during the recovery period. Among the four Census regions, the West had the largest concentration of foreclosures and the greatest volatility in the homeownership rate during 2006-2016.

3 Data

The U.S. Census Bureau reports the median housing prices and other characteristics as part of the Housing Vacancy Survey (HVS), Current Population Survey (CPS) and the American Community Survey (ACS). While Census data has the broadest coverage, it reports only median prices for owner occupied homes. For this study, I use the aggregate 5-year data from 2011 and 2016 ACS at the zip code level. These contain 'period' estimates that represent data collected over the time period 2007-2011 and 2012-2016, respectively. The primary advantage of using multiyear estimates is the increased statistical reliability of the data for less populated areas and small population subgroups. ACS 5-year estimates were first released for the period 2005-2009. The 5-year ACS is unique in its ability to produce estimates on 35 housing variables amongst other key characteristics at extremely low geographic levels.

My choice of time periods is based on the developments of the housing bubble and the resulting credit crisis. Home prices reached their peak in the second quarter of 2006. The National Bureau of Economic Research (NBER) declared the U.S. to have entered in a recession beginning December 2007 based on key economic indicators[2]. The housing bubble burst, and the U.S. housing market did not show signs of recovery until five years later (Holt (2009)). The first time period of 2007-2011 is the bust phase in the housing market and the second period of 2012-2016 can be thought of

[2]https://blogs.wsj.com/economics/2008/12/01/nber-makes-it-official-recession-started-in-december-2007

as the phase of initial recovery.

The unit of observation in this study is a Zip Code Tabulation Area (ZCTA). These were created by the Census Bureau and are generalized representations of zip codes that have been assigned to census blocks. Therefore, ZCTAs are representative of geographic locations of populated areas. I selected ZCTAs that carry the full five digit zip code and assigned region membership to these using the state Federal Information Processing Standards (FIPS) codes.

The Census Bureau has defined a system of state groupings called 'regions'[3] based on contiguous counties with the primary purpose of classification for objective statistical analysis. Regions combine all the nation's counties and statistically equivalent entities into a more concise set of general-purpose areas. The four regions have very distinct geographical, climatic, demographic, and economic characteristics and hence, characteristically different housing markets. Socioeconomic heterogeneity was the principal criterion for grouping states into regions.

Finally, I include only those ZCTAs which are present in both the time periods. As a result, I track 29,575 observations for each period. Of these, 9839, 5848, 8773, 5115 observations belong to the Midwest, Northeast, South and West regions, respectively. These comprise of both urban and rural zip codes. Figure 2 displays the nationwide and regional homeownership rates for 1990:Q1 - 2017:Q1. Homeownership rate is defined as the proportion of households that are owners. Midwest and South regions have always had higher ownership rates than the other regions and higher than the national average. During this period, homeownership reached its peak in 2005:Q1 and has been declining since then, the decline being most steep in the West.

I also collect data on the age and education of the household-head, household type, median home value, median rent, median household income and number of housing units in structure (definitions in Appendix A) to construct the variables used in my study.

[3] Appendix C

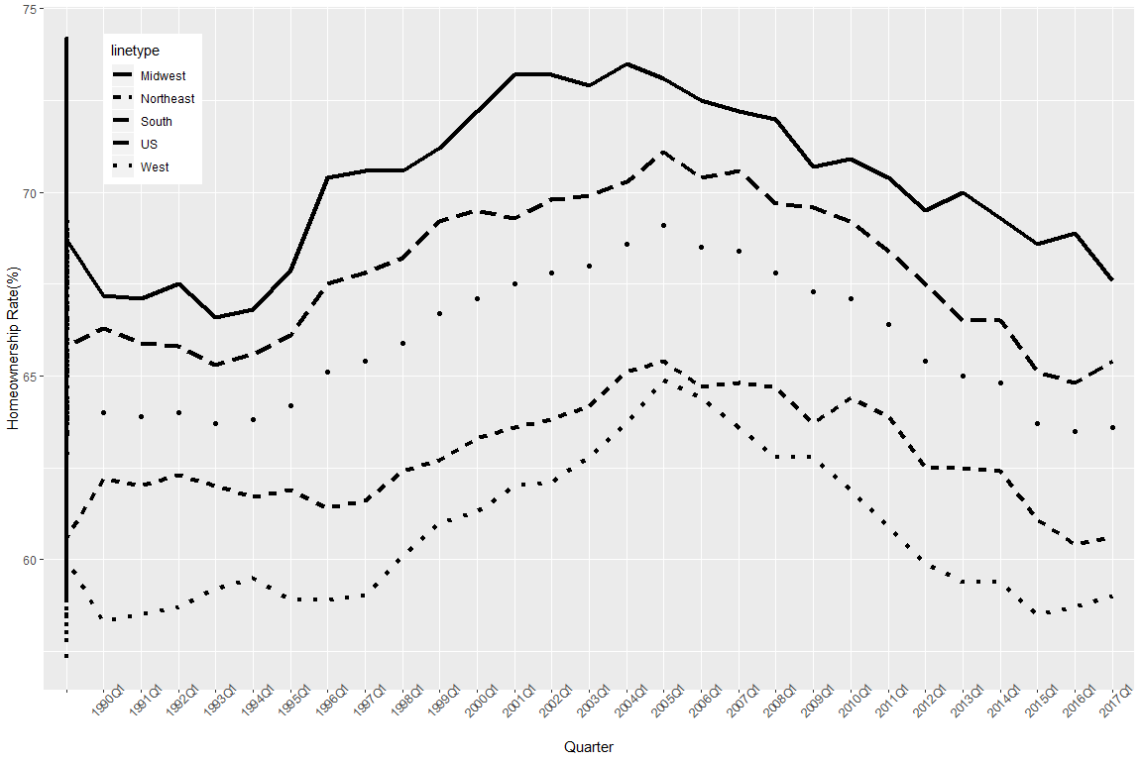


Figure 2: **Homeownership Rate for U.S. and Regions , 1990 - 2017**

Source: https://fred.stlouisfed.org/

4 Econometric Model

My regression approach is similar to Agarwal (2020), who studies cross-sectional variation in homeownership rates and develops a metric to measure unexplained variation across regions. I use six demographic and household characteristics and three dummy variables to explain variation in homeownership rates across regions and around the housing crisis. Three dummy variables are used in a linear regression model with intercept to avoid dummy variable trap with four regions. I choose 'West' as the omitted category since it has the lowest homeownership rate throughout my time period of interest, 2007-2016 (Figure 2).

The dependent variable is the percentage of occupied housing units that were

owner occupied in each of the ZCTAs, called the homeownership rate. HOR_i denotes the homeownership rate in i^{th} ZCTA. The independent variables are:

x_1 - *hhUnder35yrs.* This represents the percentage of household heads in a ZCTA that are younger than 35 years of age.

x_2 - *priceRentRatio.* This variable is computed as the ratio of median home value in a ZCTA to median annual rent.

x_3 - *priceIncRatio.* This variable is computed as the ratio of median home value in a ZCTA to median household income.

x_4 - *propNonFamily.* This measures the percentage of non-family households in a ZCTA.

x_5 - *moreThan5Units.* This measures the percentage of housing structures with five or more units.

x_6 - *bachelors.* This is the percentage of household heads with a Bachelors' degree or higher.

d_1, d_2, d_3 - *Census regional location of the ZCTA, where*
$d_1 = 1$ if ZCTA is in Midwest, 0 otherwise.
$d_2 = 1$ if ZCTA is in Northeast, 0 otherwise.
$d_3 = 1$ if ZCTA is in South, 0 otherwise.
$d_1 = d_2 = d_3 = 0$ if ZCTA is in West.

Firstly, I use the following regression model to analyze structural changes in housing market around the 2007-08 crisis.

$$\begin{aligned} HOR_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \gamma_3 d_{3i} \\ + \beta_{31} x_{3i} d_{1i} + \beta_{32} x_{3i} d_{2i} + \beta_{33} x_{3i} d_{3i} \\ + \beta_{41} x_{4i} d_{1i} + \beta_{42} x_{4i} d_{2i} + \beta_{43} x_{4i} d_{3i} \end{aligned} \tag{1}$$

where,
i is the identifier for a ZCTA,
β_j is the slope coefficient for variable j for West, and
$(\beta_j + \beta_{jk})$ is the slope coefficient for variable j for,

Midwest if $k = 1$,

Northeast if $k = 2$,

South if $k = 3$.

γ_k is the slope coefficient for dummy variable k.

Equation (1) is evaluated for both time periods and the coefficients are tested for joint equality using the Chow test (Chow (1960)). The Chow test is a test of whether the true coefficients in two linear regressions on different data sets are equal. The null hypothesis of the Chow test asserts that,

$$\beta_i^{T_1} = \beta_i^{T_2} \quad \forall \quad i$$

where $\beta_i^{T_1}$ denotes regression coefficients from the first time period (2007-2011) and $\beta_i^{T_2}$ denotes regression coefficients from the second time period (2012-2016).

The Chow test statistic is given by $\mathcal{F}_{stat}$ and follows the $\mathcal{F}$-distribution.

$$\mathcal{F}_{stat} = \frac{(S_c - (S_1 + S_2))/k}{(S_1 + S_2)/(N_1 + N_2 - 2k)} \sim \mathcal{F}_{(k, N_1+N_2-2k)}$$

where,
S_c is the sum of squared residuals from the combined data,
S_i is the sum of squared residuals from the i^{th} group for $i = 1, 2$,
N_i is the number of observations in the i^{th} group for $i = 1, 2$, and
k is the total number of parameters.

Next, I compare the region-specific slope coefficients of the variable $priceIncRatio$ from both time periods to test my hypothesis regarding the change in homeownership response as a result of changed lending practices.

Finally, similar to Agarwal (2020), I also evaluate the following linear models for each time period,

Model A:

$$HOR_i = \gamma_0 + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \gamma_3 d_{3i}$$

Model B:

$$HOR_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \gamma_3 d_{3i}$$

Model A is a baseline model with the homeownership rates regressed only on the regional dummy variables with *West* as the omitted category. Model B includes all the covariates without any interaction terms so as to focus only on the regional variation.

To quantify the variability into a single measure, I define the sum of squared differences (SSD) as the sum of squares of pairwise differences between regression coefficients on the dummy variables,

$$SSD = \sum_{i=0}^{2} \sum_{j=i+1}^{3} (d_i - d_j)^2$$

where d_0, d_1, d_2 and d_3 indicate West, Midwest, Northeast and South respectively. Next, I compute the percentage reduction in the SSD between models A and B as below,

$$\Delta = \frac{SSD^B - SSD^A}{SSD^A} * 100$$

This measure, Δ is the relevant test statistic. I generate 5000 bootstrap samples with replacement to get a distribution of Δ for each of the two time periods (Hansen (2020)). Next, I utilize a paired t-test of significance to test for statistically significant differences in explanatory power of variables across time.

5 Results

Table 1 shows the results for linear regression model outlined in equation (1) for time periods, T_1 (2007-2011) and T_2 (2012-2016). Values indicate the slope coefficients for

Table 1: **Regression Results**

Variable	(1) T_1 : 2007-2011	(2) T_2 : 2012-2016
x_0 - intercept	89.0050***	86.1003***
x_1 - hhUnder35yrs	-0.5190***	-0.6103***
x_2 - priceRentRatio	-0.0019	0.0787*
x_3 - priceIncRatio	-1.0609***	-1.3739***
x_4 - propNonFamily	-0.2398***	-0.1767***
x_5 - moreThan5Units	-0.5071***	-0.5241***
x_6 - bachelors	0.2444***	0.1995***
d_1 - mw	3.9221***	4.6340***
d_2 - ne	5.2992***	5.3437***
d_3 - so	1.4489*	0.8933
x_2d_1 - priceRentRatio:mw	0.2545***	0.4271***
x_2d_2 - priceRentRatio:ne	0.1007***	0.1262**
x_2d_3 - priceRentRatio:so	0.0678	0.2643***
x_3d_1 - priceIncRatio:mw	-0.7948**	-1.8592***
x_3d_2 - priceIncRatio:ne	-1.0779***	-1.1279***
x_3d_3 - priceIncRatio:so	0.0490	-0.7590(.)
adj-R^2	0.6860	0.7304
AIC	185297	173746
BIC	185436	173884

Note 1: The table shows the results of a regression investigating the relationship between various explanatory factors and homeownership rate in a ZCTA.

Note 2: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 respectively.

Note 3: Significance levels indicate $Pr(> |t|)$ using White standard errors.

each variable and significance levels indicate $Pr(>|t|)$ using White robust standard errors for heteroscedasticity. Table 2 shows three types of standard errors for all independent variables included in the model for both time periods. Columns (1) and (4) show the ordinary least squares (OLS) standard errors. OLS estimators are not the best linear unbiased estimators in the presence of heteroscedasticity in the class of linear unbiased estimators. As a result, they may lead to biased inference. White robust standard errors are also called heteroscedasticity-consistent (HC) standard errors. White standard errors are larger than OLS standard errors for both T_1 and T_2 for all coefficients. However, most coefficients are statistically significant at 0.0001 levels based on either OLS standard errors or White standard errors. White standard errors result in lower significance than OLS standard errors for three coefficients (*priceRentRatio*, *so* and *priceRentRatio:so*) for T_1 and four coefficients (*priceRentRatio*, *so*, *priceRentRatio:ne* and *priceIncRatio:so*) for T_2.

Table 2 also reports another type of standard errors in columns (3) and (6). Clustered standard errors may be considered as extensions to HC standard errors. These assume that the errors are uncorrelated across clusters while errors for individuals belonging to the same cluster may be correlated (Abadie et al. (2017), Cameron and Miller (2015)). In my study, observations belong to four census regions which can be thought of as clusters. Comparing columns (3) and (6) with (1) and (4) respectively, clustered standard errors are generally larger than the OLS standard errors for all quantitative variables for both time periods. However, this is not the case for dummy variables pertaining to regions and the interaction variables. All coefficients have the same level of statistical significance based on clustered standard errors and OLS standard errors (except dummy variable, *so* in T_1).

5.1 Analysis of Hypothesis H1a

Table 1 shows the regression results. The same regression model is evaluated for a consistent sample of ZCTAs over T_1 and T_2. Intercept represents mean conditional homeownership rate for the West. Regional dummy variables indicate marginal effects on homeownership rate of a ZCTA located in a region relative to the West.

Table 2: **OLS, White Robust, and Clustered Standard Errors**

Variable	(1) OLS se	(2) White se	(3) Clustered se	(4) OLS se	(5) White's se	(6) Clustered se
		T_1 : 2007-2011			T_2 : 2012-2016	
x_0 - intercept	0.3359***	0.5352***	0.7471***	0.3381***	0.5182***	1.0633***
x_1 - hhUnder35yrs	0.0072***	0.0128***	0.0316***	0.0077***	0.0132***	0.0468***
x_2 - priceRentRatio	0.0084	0.0121	0.0033	0.0169***	0.0385*	0.0201***
x_3 - priceIncRatio	0.0522***	0.0904***	0.0863***	0.0731***	0.1860***	0.0764***
x_4 - propNonFamily	0.0059***	0.0089***	0.0197***	0.0060***	0.0084***	0.0258***
x_5 - moreThan5Units	0.0059***	0.0093***	0.0169***	0.0058***	0.0096***	0.0197***
x_6 - bachelors	0.0046***	0.0063***	0.0229***	0.0041***	0.0048***	0.0270***
d_1 - mw	0.3717***	0.7068***	0.1488***	0.3867***	0.5844***	0.1352***
d_2 - ne	0.3748***	0.5610***	0.0865***	0.3793***	0.5561***	0.1929***
d_3 - so	0.3460***	0.6194*	0.1590***	0.3661*	0.6385	0.2479***
x_2d_1 - priceRentRatio:mw	0.0172***	0.0589***	0.0081***	0.0260***	0.0468***	0.0142***
x_2d_2 - priceRentRatio:ne	0.0164***	0.0237***	0.0153***	0.0241***	0.0451**	0.0112***
x_2d_3 - priceRentRatio:so	0.0162***	0.0488	0.0089***	0.0269***	0.0672***	0.0084***
x_3d_1 - priceIncRatio:mw	0.1203***	0.2601**	0.0854***	0.1531***	0.2655***	0.1751***
x_3d_2 - priceIncRatio:ne	0.0847***	0.1370***	0.0313***	0.1061***	0.2180***	0.0247***
x_3d_3 - priceIncRatio:so	0.0965	0.2606	0.0615	0.1293***	0.4489(.)	0.1179***

Note 1: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Note 2: White standard errors represent heteroscedasticity-consistent (HC) standard errors.

Note 3: Clustered standard errors assume that the errors are uncorrelated across regions while errors for individual ZCTAs belonging to the same region may be correlated.

All the demographic variables - $hhUnder35yrs$, $propNonFamily$ and $bachelors$ have the expected signs and high statistical significance for both time periods. Age and education can be considered as proxies for permanent income and hence are expected to correlate positively to homeownership (Coulson (2002)). Non-family households consist of people who live alone or share their residence with unrelated individuals. Individuals are more comfortable in committing to a mortgage with a spouse than individually or with an unrelated person. As a result, households with married couples are more likely to be homeowners. Thus, $propNonFamily$ has the expected negative sign.

Hypothesis H1a predicts that the relationship between zip code level homeownership rate and its determinants changed in the recovery period (2012-2016) compared to the period of the housing crisis (2007-2011) due to changes in the household behavior and housing market dynamics. The adjusted-R^2 for T_1 is 68.6% and for T_2 is 73%. This implies that the factors that can explain 73% variation in homeownership rate across zip codes in recovery phase, explain only 68.6% variation in the period of the housing crash. This is indicative of a structural change in the housing market and demand for homeownership.

Amongst the quantitative variables, all coefficients except for $propNonFamily$ and $bachelors$, increased in magnitude. Conditional on all other included variables, if the $hhUnder35yrs$ in a zip code is higher by one percentage point compared to another similar zip code, then irrespective of the regional location of the zip code, it is expected to have a lower homeownership rate by approximately 0.5 percentage points in T_1 and by 0.6 percentage points in T_2.

Relationship between homeownership and market variables also seems to have changed significantly for all regions. The coefficient on $priceRentRatio$ which indicates the effect in the West, changed from -0.0019 (not significantly different from zero) in T_1 to 0.079 with somewhat higher significance in T_2. The marginal effects of region on $priceRentRatio$ also increased sharply for all regions. The most striking of these are the coefficients for the South. From column (1) in table 1, $priceRentRatio:so$ is just 0.0678 which is not significantly different relative to the West. However, for T_2 (column (2)), the coefficient on $priceRentRatio$ for the South is higher than

the West by 0.2643 and is significant at 0.001 level of significance.

I utilize the Chow test to test for joint equality of regression coefficients across the two time periods. It is the seminal test to detect a structural break or shift in parameters of the relevant model due to some external shock. The null hypothesis of a Chow test asserts that the underlying model yields better explanatory power when fit to the combined data instead of grouped data.

I find that Chow test rejects the null hypothesis of no structural break at 0.001 level of significance. This implies that the model fit on two groups of data provides better explanatory power than the model fit to the combined data set. My results conclude that there was a significant effect of the bubble on the cross-sectional homeownership behavior at the zip code level ($\mathcal{F}_{(16,29543)} = 70.27$, $p < .00001$).

5.2 Analysis of Hypothesis H1b

Regional dummy variables in the regression equation account for the regional variability in homeownership rates. This is an estimate of the difference in homeownership rates, conditional on covariates included in the model, i.e., how much of the regional differentiation is due to unobservables. I use a naive approach to build and test hypothesis (H1b). Table 3 shows the results of my exercise.

Model A is a baseline model with the homeownership rates regressed only on the regional dummy variables with *West* as the omitted category. Model B includes all the covariates without any interaction terms so as to focus only on the regional variation. All the variables are highly significant and have the expected signs for both time periods. I define the Sum of Squared Differences (SSD) as a measure of total inter-regional variation that cannot be explained by the explanatory variables in my model. SSD is the sum of squares of pairwise difference between regression coefficients of the dummy variables. There are six unique pairwise differences for four regions. For instance, SSD for Model 1A is equal to 209.63 while that for Model 2A is equal to 214.99.

The SSD by itself does not convey any evidence about regional variation across time periods. To address this, I define the measure Δ which is the percentage

Table 3: **Sum of Squared Differences**

	T_1 : 2007-2011		T_2 : 2012-2016	
Variable	Model 1A	Model 1B	Model 2A	Model 2B
x_0 - intercept	68.5325***	88.7356***	67.1214***	86.1904***
x_1 - hhUnder35yrs	—	-0.5204***	—	-0.6094***
x_2 - priceRentRatio	—	0.0668***	—	0.2543***
x_3 - priceIncRatio	—	-1.4442***	—	-2.1514***
x_4 - propNonFamily	—	-0.2387***	—	-0.1805***
x_5 - moreThan5Units	—	-0.5131***	—	-0.5284***
x_6 - bachelors	—	0.2542***	—	0.2025***
d_1 - mw	9.8455***	6.0484***	9.8332***	5.8006***
d_2 - ne	7.3080***	3.2837***	7.5621***	2.9862***
d_3 - so	5.2749***	3.1926***	4.8079***	2.5226***
SSD	209.63	73.37	214.99	67.81
Δ	-65.0%		-68.5%	

Note 1: Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Note 2: Sum of Squared Differences (SSD) is defined as the sum of squares of pairwise differences between regression coefficients of the dummy variables,

$$SSD = \sum_{i=0}^{2} \sum_{j=i+1}^{3} (d_i - d_j)^2$$

where d_0, d_1, d_2 and d_3 indicate West, Midwest, Northeast and South respectively. $d_0 = 0$ and d_1, d_2, d_3 are the regression coefficients for mw, ne and so respectively.

Note 3: Delta (Δ) is defined as the percentage reduction in the SSD by including the explanatory variables,

$$\Delta = \frac{SSD^B - SSD^A}{SSD^A} * 100$$

Note 4: For four regions, there are six distinct absolute pairwise differences which are $|d_0 - d_1|$, $|d_0 - d_2|$, $|d_0 - d_3|$, $|d_1 - d_2|$, $|d_1 - d_3|$ and $|d_2 - d_3|$.

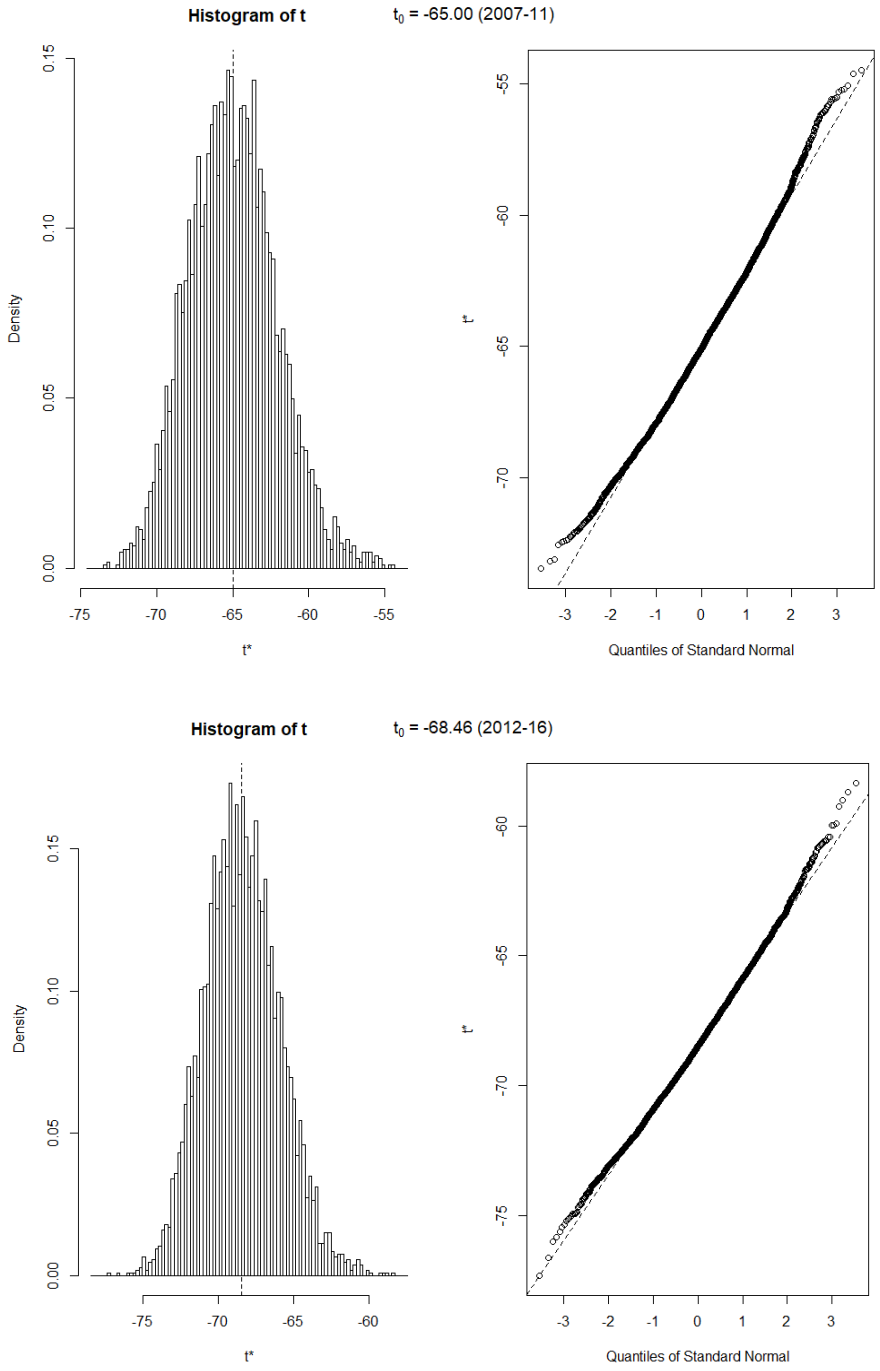


Figure 3: **Bootstrap Distribution for Test-statistic, Δ**

Note:Number of bootstrap samples, $B = 5000$

reduction in the SSD by including the explanatory variables in the model. This step yields a single measure for each time period. The Δ for T_1 is -65% and for T_2 is -68.5%. The point estimates of Δ indicate an improvement of only 3.5% in the reduction in regional differentiation due to unobserved factors.

In order to test for the statistical significance, I create a bootstrapped distribution of the statistic, Δ with replacement and then perform a paired t-test. Bootstrapped distribution is approximately normal (see Figure 3) for both time periods and is centered around the observed point estimate. I check for stability of the test-statistic and bootstrap algorithm by iterating over different random seeds and different bootstrap sample sizes (Appendix B). Δ follows a $t-$distribution with 4999 degrees of freedom.

A paired $t-$test on the difference of means ($t_{4999} = 3.41867$, $p < 2.2e-16$) concludes that the regional variation due to unobserved factors has indeed decreased during the recovery period. In other words, housing market fundamentals have a more homogeneous effect on homeownership rate across regions. This is consistent with the prediction of hypothesis H1b.

5.3 Analysis of Hypothesis H2

Hypothesis H2 predicts that the homeownership rate became more strongly related to the ratio of median home value to median household income in the recovery phase (T_2) compared to the bust phase (T_1). I focus on the coefficients for *priceIncRatio* and its interaction with the regional dummy variables from Table 1. The region *West* is omitted for the region dummy variables. Regression coefficient on *priceIncRatio* indicates the change in the homeownership rate in percentage points for every unit change in the *priceIncRatio* of a ZCTA for the West region, conditional on other explanatory variables. The coefficient on the interaction of *priceIncRatio* should be interpreted relative to the West.

The median household income can be considered a good proxy for permanent income in a ZCTA (Eilbott and Binkowski (1985)). The demand for ownership increases with rising income, and higher income makes it easier to obtain a mortgage.

Table 4: **Net Effect of Variables**

Variable(x_i)		(1) β_i	(2) σ_x	(3) β_i^*	(4) β_i	(5) σ_x	(6) β_i^*
		T_1 : 2007-2011			T_2 : 2012-2016		
hhUnder35yrs		-0.5190	10.1	-5.2419	-0.6103	9.5	-5.7979
priceRentRatio	*West*	-0.0019	18.4	-0.0350	0.0787	12.4	0.9759
	Midwest	0.2526	7.3	1.8440	0.5058	5.6	2.8325
	Northeast	0.0988	11.3	1.1164	0.2049	9.3	1.9056
	South	0.0659	9.2	0.6063	0.3430	6.4	2.1952
priceIncRatio	*West*	-1.0609	7.1	-7.5324	-1.3739	2.8	-3.8469
	Midwest	-1.8557	1.6	-2.9691	-3.2331	0.9	-2.9098
	Northeast	-2.1388	2.5	-5.3470	-2.5018	2.1	-5.2538
	South	-1.0119	2.4	-2.4286	-2.1329	1.4	-2.9861
propNonFamily		-0.2398	12.7	-3.0455	-0.1767	12.3	-2.1734
moreThan5Units		-0.5071	13.3	-6.7444	-0.5241	13.5	-7.0754
bachelors		0.2444	14.7	3.5927	0.1995	17.3	3.4514

Note: The table shows the values of region-specific standardized and unstandardized regression coefficients. β_i represents percentage point change in homeownership rate for a 1 unit change in the value of i^{th} variable. β_i^* represents percentage point change in homeownership rate for a 1 standard deviation change in the value of i^{th} variable. σ_i is the standard deviation of i^{th} variable.

$$\beta_i^* = \beta_i * \sigma_i$$

At the same time, the demand for homeownership decreases with rising home prices in an area. Thus, the ratio of median home value to the median household income serves as a measure of housing affordability in an area and is expected to have a negative impact on homeownership.

For T_1 in column (1) of Table 1, the homeownership rate decreases by a 1.06 percentage points for every unit increase in *priceIncRatio* for a ZCTA in the West. Similarly for T_2 in column (2), the homeownership rate decreases by 1.37 percentage points for every unit increase in *priceIncRatio* for a ZCTA in the West. Coefficients for both time periods are highly significant and have the expected signs for the West.

The effect of *priceIncRatio* in the other three regions in T_1 is also shown in column (1) of Table 1. Even after adjusting for the other factors included in the regression, the absolute magnitude of effect of *priceIncRatio* on homeownership rate is 0.8 percentage points higher in the Midwest than the West, 1.1 percentage points higher in the Northeast than the West and not significantly different in the South than the West.

Column (2) of Table 1 reports these effects for T_2. Conditional on the other included variables, the effect of *priceIncRatio* on homeownership in absolute terms is 1.9 percentage points higher in the Midwest than the West and 1.1 percentage points higher in the Northeast than the West. Both of these coefficients are statistically significant at $\alpha = 0.001$. However, responsiveness of homeownership to *priceIncRatio* in absolute terms is 0.8 percentage points higher in the South than in the West, which is significant only at the 10% level.

These results are consistent with hypothesis H2. While regression coefficients for *priceIncRatio* increased for all regions in $T2$, the marginal effects for zip codes in the Midwest, Northeast and South also increased relative to the West during the recovery period. Coefficient for the South is not statistically different from the West since both of these regions had the largest concentration of foreclosures and were most affected by the housing crisis. Home price appreciation was generally more modest in the Midwest and Northeast regions. In fact, housing market in the Northeast was the least volatile among the four Census regions (Rosen et al. (2017)).

Next I focus on the net coefficients for variable *priceIncRatio* for each region as

shown in Table 4. Columns (1) and (4) report the net regression coefficients for T_1 and T_2. Overall the net effect of an increase in the ratio of median home value to median household income in a ZCTA is larger in T_2, the initial recovery phase than the T_1, across all four regions. For instance, during the bust phase, an increase of 1 unit in the *priceIncRatio* of a ZCTA in the Midwest would be expected to decrease the homeownership rate there by less than 2 percentage points. This is because a bubble formed in the housing markets as home prices across the country increased each year from the mid 1990s to 2006, moving out of line with fundamentals like household income (Baily et al. (2008)). As credit mortgage defaults escalated, credit availability decreased and mortgage underwriting standards tightened. As can be seen from column (4) in Table 4, in the initial recovery phase, a unit increase in the *priceIncRatio* for a ZCTA in the Midwest predicts a decrease of more than 3 percentage points in the homeownership rate. The results are similar across all regions.

Table 4 also shows the standardized effects of a unit standard deviation change in the right-hand side variable in columns (3) and (6). β^* represents the percentage point change in the homeownership rate for one standard deviation change in the magnitude of the explanatory variable. Columns (2) and (4) report the standard deviation of the respective variables in my sample dataset.

It appears that the spread of the distribution of ratio of median home values to median household incomes reduced significantly during the recovery period. This occurrence leads to similar region-specific homeownership responses before and after the bursting of housing bubble, with the exception of the West. It is also worth pointing out from columns (2) and (4) in Table 4, the standard deviation of *priceIncRatio* decreased the most for the West and the least for the Northeast in T_2 relative to T_1.

For zip codes in the West, a *priceIncRatio* higher by one standard deviation would, on average, have a homeownership rate lower by 7.5 percentage points compared to a similar zip code in T_1. In T_2, due to a sharp decrease in standard deviation of *priceIncRatio* from 7.1 to 2.8, the magnitude of standardized coefficient is only 3.85 for the West. For the Midwest, the increase in magnitude of slope coefficient and the decrease in standard deviation seem to nearly offset the opposing effects. If

the *priceIncRatio* in a zip code in the Midwest is higher than that in another similar zip code by one standard deviation, then homeownership rate in the zip code with higher *priceIncRatio* is expected to be lower by 2.97 percentage points in T_1 and by 2.91 percentage points in T_2.

6 Conclusion

This paper investigates the impact of the housing bubble on homeownership during the housing bubble bust (2007-2011) and the initial phase of economic recovery (2012-2016) both at the national and the regional levels. Since a major part of the run-up in homeownership was due to relaxed credit standards (Goodman and Mayer (2018)), I also investigate how has the homeownership response to relative affordability changed before and after the crash of house prices.

I use a sample of 29,575 zip codes from the aggregate 5-year data from 2011 and 2016 American Community Survey (ACS) at the zip code level. The first time period of 2007-2011 is the bust phase in the housing market and the second period of 2012-2016 can be thought of as the phase of initial recovery. I construct six key demographic and economic variables for each zip code from the ACS data and assign the regional membership to each of these.

The key contributions of this study are threefold. Firstly, existing studies are based on a data from individual households or on aggregate data for only Metropolitan Statistical Areas (MSA). No investigation of the ownership outcomes across regions and across time has used aggregate data for a large number of zip codes. This paper uses data for almost all zip codes covered by the 5-year American Community Survey (ACS). This enables me to account for heterogeneity across submarkets within MSAs, not just across MSAs. Secondly, the issue of regional disparity has largely been discussed using data from the cities or the metro areas and the rural areas have been left out of such studies. In fact, Glaeser and Gyourko (2002) point out that smaller datasets which feature only large metropolitan areas may oversample highly regulated cities and underrepresent the bulk of American housing. Lastly, I create a naive metric and test statistic to confirm the regime change in the housing

market at the regional level.

The region can affect the homeownership rate of an area not only through its unobserved locational effect but also through interaction with demographic and market variables. I use six explanatory variables, three dummy variables and relevant interaction terms to explain variation in homeownership rates across regions and around the housing crisis. I find that the housing markets were indeed very different before and after the housing crash of 2007-08 at both the national and the regional level. I test this using Chow test for structural break and find sufficient evidence to reject the null hypothesis of no structural break in the homeownership model. Based on adjusted-R^2, the regression model has a higher explanatory power in the recovery period compared to the crisis period.

I look further into it by using dummy coefficients to estimate the regional variation in homeownership rates. I find that fundamental demographic and market variables reduce the collective variation more in the recovery period than in the crisis period. This reduction is also statistically significantly different. Using a bootstrapped distribution of this statistic, I find that my included factors have more homogeneous effect on homeownership rate across regions in the recovery period than in the crisis period.

Next, I focus only on the response of homeownership rate in a zip code to the ratio of median home value to median household income. Since there was a dramatic shift form lax lending standards to a much tighter credit regime, I expect the homeownership to be more responsive in the recovery period. I find that an increase in this affordability measure leads to a much larger decrease in homeownership rate during the recovery phase compared to the response in the bust phase. This result is also statistically significant and consistent for all regions.

My study presents opportunities for future research in two dimensions. Motivated by expectation of a regime switch before and after the housing market recovery, I focus on homeownership analysis across only two time periods. Extending the analysis to a longer time horizon would shed light on the homeownership behavior during the economic boom and house price appreciation that occurred prior to the crash (2002-2007). However, insufficient history of the ACS 5-year aggregate data

limits the prospects of conducting a full panel analysis here. Since these estimates were first released for the period 2005-2009, only three distinct, non-overlapping time periods are available to date. Other data sources may be considered that provide estimates over a longer time period to conduct a panel data study.

Secondly, beginning in the late 1970s, state and local governments, have become increasingly important to the development and implementation of housing policy (Buckley and Schwartz (2011)). My study outlines a framework to analyzes regional variation in homeownership rates by insertion of dummy variables and interactions with market variables. While this uncovers interesting insights into key determining factors, this obscures the role of factors contributing to inter-state variations in homeownership rates.

Aside from factors that are commonly associated with housing demand (demographics, income and density), housing prices are found to be associated with local and statewide regulations (Eicher (2008)). State policies in the form of land use regulations, zoning restrictions and construction approval delays may add a complementary supply-side dimension of the variation in housing prices to the analysis. The Wharton Residential Land Use Regulatory Index (WRLURI) is an aggregate measure that summarizes information on the different aspects of local and state regulatory environment (Gyourko et al. (2008)).

Stricter land use regulations tend to stifle or delay new supply, leading to a rise in land prices and construction costs, and a decrease in homeownership (Rosen et al. (2017)). Additionally, past studies have also identified differential state income taxes leading to differences in homeownership. For instance, Narwold and Sonstelie (1994) show that the probability that a household owns its home is positively related to its own tax rate and is negatively related to the top tax rate in its state. Thus, explaining state level variation in homeownership is a promising topic for future study.

References

Aalbers, M. (2009). Geographies of the financial crisis. *Area 41*(1), 34–42.

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Acolin, A., J. Bricker, P. Calem, and S. Wachter (2016). Borrowing constraints and homeownership. *American Economic Review 106*(5), 625–29.

Agarwal, S. (2020). *Spatial Heterogeneity in Homeownership Rates*. Ph. D. thesis, Seattle, WA.

Ashcraft, A. B. and T. Schuermann (2008). *Understanding the securitization of subprime mortgage credit*. Number 318. Now Publishers Inc.

Baily, M. N., R. E. Litan, M. S. Johnson, et al. (2008). The origins of the financial crisis.

Barth, J. R., G. Caprio, and R. Levine (2013). Bank regulation and supervision in 180 countries from 1999 to 2011. *Journal of Financial Economic Policy*.

Bhattacharya, R. and S.-W. Kim (2011). Economic fundamentals, subprime lending and housing prices: Evidence from msa-level panel data. *Housing Studies 26*(6), 897–910.

Bricker, J., B. Bucks, A. B. Kennickell, T. Mach, and K. B. Moore (2011). Surveying the aftermath of the storm: Changes in family finances from 2007 to 2009.

Brueckner, J. K., P. S. Calem, and L. I. Nakamura (2012). Subprime mortgages and the housing bubble. *Journal of Urban Economics 71*(2), 230–243.

Buckley, R. M. and A. F. Schwartz (2011). Housing policy in the us: the evolving sub-national role. *International Affairs Working Paper 6*.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of human resources 50*(2), 317–372.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, 591–605.

Clark, S. P. and T. D. Coggin (2009). Trends, cycles and convergence in us regional house prices. *The Journal of Real Estate Finance and Economics 39*(3), 264–283.

Coulson, N. E. (2002). Regional and state variation in homeownership rates; or if california's home prices were as low as pennsylvania's would its ownership rate be as high? *The Journal of Real Estate Finance and Economics 24*(3), 261–276.

Cukierman, A. (2013). Monetary policy and institutions before, during, and after the global financial crisis. *Journal of Financial Stability 9*(3), 373–384.

De Nardi, M., E. French, and D. Benson (2011). Consumption and the great recession. Technical report, National Bureau of economic research.

Dietz, R. D. and D. R. Haurin (2003). The social and private micro-level consequences of homeownership. *Journal of urban Economics 54*(3), 401–450.

DiPasquale, D. and E. Glaeser (1999). Incentives and social capital: Are homeowners better citizens?

Eicher, T. S. (2008). Housing prices and land use regulations: A study of 250 major us cities. *Northwest Journal of Business and Economics 2008*.

Eilbott, P. and E. S. Binkowski (1985). The determinants of smsa homeownership rates. *Journal of Urban Economics 17*(3), 293–304.

Elsinga, M. and J. Hoekstra (2005). Homeownership and housing satisfaction. *Journal of Housing and the Built Environment 20*(4), 401–424.

Glaeser, E. L. and J. Gyourko (2002). The impact of zoning on housing affordability. Technical report, National Bureau of Economic Research.

Glumov, D. (2013). How did the great financial crisis affect portfolio allocations and attitudes towards risk? evidence from the'survey of consumer finances'. *Evidence from the'Survey of Consumer Finances'(July 12, 2013)*.

Goodman, L., J. Zhu, and T. George (2015). Four million mortgage loans missing from 2009 to 2013 due to tight credit standards. *Urban Institute*.

Goodman, L. S. (2018). What is holding back housing? *Business Economics 53*(2), 79–85.

Goodman, L. S. and C. Mayer (2018). Homeownership and the american dream. *Journal of Economic Perspectives 32*(1), 31–58.

Gwartney, J., D. Macpherson, R. Sobel, and R. Stroup (2008). The crash of 2008: Cause and aftermath. *Cencage. com. http://www. cengage. com/economics/boo k_ content/0324580185_ gwartney/conten tihtml*.

Gyourko, J., A. Saiz, and A. Summers (2008). A new measure of the local regulatory environment for housing markets: The wharton residential land use regulatory index. *Urban Studies 45*(3), 693–729.

Hansen, B. E. (2020, June). Bruce hansen's econometrics. https://www.ssc.wisc.edu/~bhansen/econometrics/. (Accessed on 07/08/2020).

Holt, J. (2009). A summary of the primary causes of the housing bubble and the resulting credit crisis: A non-technical paper. *The Journal of Business Inquiry 8*(1), 120–129.

Institute, U. (2016, November). Housing finance at a glance. https://www.fhfa.gov/SupervisionRegulation/FannieMaeandFreddieMac/Pages/About-Fannie-Mae---Freddie-Mac.aspx. (Accessed on 07/11/2020).

Mattews, W. and D. Robert (2016). The global financial crisis and us housing policy. *Journal of Business and Retail Management Research (JBRMR)*.

Mian, A., K. Rao, and A. Sufi (2013). Household balance sheets, consumption, and the economic slump. *The Quarterly Journal of Economics 128*(4), 1687–1726.

Miles, W. (2015). Regional house price segmentation and convergence in the us: A new approach. *The Journal of Real Estate Finance and Economics 50*(1), 113–128.

Narwold, A. and J. Sonstelie (1994). State income taxes and homeownership: a test of the tax arbitrage theory. *Journal of Urban Economics 36*(3), 249–277.

Nissan, E. and J. E. Payne (2013). A simple test of σ-convergence in us housing prices across bea regions. *Journal of Regional Analysis and Policy 43*(1100-2016-90116), 178–185.

Nneji, O., C. Brooks, and C. W. Ward (2013). House price dynamics and their reaction to macroeconomic changes. *Economic Modelling 32*, 172–178.

O'Neill, B. and J. J. Xiao (2012). Financial behaviors before and after the financial crisis: Evidence from an online survey. *Journal of Financial Counseling and Planning 23*(1).

Ozdemir, D. (2020). Time-varying housing market fluctuations: Evidence from the us housing market. *Real Estate Management and Valuation 28*(2), 89–99.

Painter, G. and C. L. Redfearn (2002). The role of interest rates in influencing long-run homeownership rates. *The Journal of Real Estate Finance and Economics 25*(2-3), 243–267.

Pavlov, A. D. and S. M. Wachter (2009). Subprime lending and house price volatility. *U of Penn, Inst for Law & Econ Research Paper* (08-33).

Petev, I., L. Pistaferri, and I. S. Eksten (2011). Consumption and the great recession: An analysis of trends, perceptions, and distributional effects.

Rosen, K., D. Bank, A. Patel, B. Fawley, H. Belfor, A. Eckstein, and A. Gade (2017). Rebuilding the american dream: Strategies to sustainably increase homeownership.

Rosen, K. T., D. Bank, H. Belfor, A. Eckstein, A. Gade, and M. Tcheau (2017). Hurdles to homeownership: Understanding the barriers.

Rosen, K. T., D. Bank, A. Eckstein, M. Stern, and M. Tcheau (2017). Homeownership in crisis: Where are we now?

Rossi, P. H. and E. Weber (1996). The social benefits of homeownership: Empirical evidence from national surveys. *Housing policy debate 7*(1), 1–35.

Schloemer, E. (2006). *Losing Ground: Foreclosures in the Subprime Market & Their Cost to Homeowners*. DIANE Publishing.

Tan, J. and S. A. Cheong (2016). The regime shift associated with the 2004–2008 us housing market bubble. *PloS one 11*(9), e0162140.

Taylor, P., R. Morin, R. Kochhar, K. Parker, D. Cohn, M. H. Lopez, R. Fry, W. Wang, G. Velasco, D. Dockterman, et al. (2010). A balance sheet at 30 months: How the great recession has changed life in america. *Washington, DC: Pew Research Center*.

Turner, T. M. and H. Luea (2009). Homeownership, wealth accumulation and income status. *Journal of Housing Economics 18*(2), 104–114.

WPR-2019-001, O. o. I. G. (2019). Subprime mortgages: Enterprise and fhfa reporting. Technical report, Federal Housing Finance Agency.

WPR-2019-002, O. o. I. G. (2019). An overview of enterprise debt-to-income ratios. Technical report, Federal Housing Finance Agency.

WSJ (2008, December). NBER makes it official: Recession started in december 2007 - real time economics - wsj. https://blogs.wsj.com/economics/2008/12/01/nber-makes-it-official-recession-started-in-december-2007. The Wall Street Journal.

A Data Definitions

Metropolitan Statistical Area (MSA). A census-designated area with at least one urbanized area of 50,000 or more inhabitants.

Micropolitan Statistical Area. A census-designated area with at least one urban cluster of at least 10,000 but less than 50,000 population.

Housing Unit. A housing unit is a house, an apartment, a group of rooms, or a single room occupied or intended for occupancy as separate living quarters. Separate living quarters are those in which the occupants do not live and eat with other persons in the structure and which have direct access from the outside of the building or through a common hall. Living quarters of the following types are excluded from the housing unit inventory: Dormitories, bunkhouses, and barracks; quarters in predominantly transient hotels, motels, and the like, except those occupied by persons who consider the hotel their usual place of residence; quarters in institutions, general hospitals, and military installations except those occupied by staff members or resident employees who have separate living arrangements.

Occupied Housing Units. A housing unit is occupied if a person or group of persons is living in it at the time of the interview or if the occupants are only temporarily absent, as for example, on vacation. The persons living in the unit must consider it their usual place of residence or have no usual place of residence elsewhere. The count of occupied housing units is the same as the count of households.

Householder. The householder refers to the person (or one of the persons) in whose name the housing unit is owned or rented or, if there is no such person, any adult member, excluding roomers, boarders, or paid employees. If the house is jointly owned by a married couple, either the husband or the wife may be listed first, thereby becoming the reference person, or householder, to whom the relationship of the other household members is recorded. One person in each household is designated as the "householder."

Number of Housing Units in Structure. A structure is a separate building which either has open space on all four sides or is separated from other structures by dividing walls that extend from ground to roof. The count of housing units in a structure is the total number of units in the structure, both occupied and vacant units. In the tabulations, occupied mobile homes or trailers, tents, and boats are included in the category one housing unit in structure.

Contract Rent. For renter-occupied units, the contract rent is the monthly rent agreed upon regardless of any furnishings, utilities, or services that may be included. The median monthly rent is the rent which divides the series into two equal parts, one-half of the units with rents higher than the median and the other half with rents lower than the median.

Value. Value is the respondent's estimate of how much the property would sell for on the current market. For vacant units, value is the sales price asked for the property at the time of the interview and may differ from the price at which the property is sold. The median value or sales price is the amount which divides the series into two equal parts, one-half of the units with values higher than the median and the other half with values lower than the median.

Homeownership Rate. The proportion of households that are owners is termed the homeownership rate. It is computed by dividing the number of households that are owners by the total number of households. The formula is as follows:

$$\textit{Homeownership Rate (\%)} = \left[\frac{\textit{Owner occupied housing units}}{\textit{Total occupied housing units}}\right] * 100$$

Family. A family is a group of two persons or more (one of whom is the householder) related by birth, marriage, or adoption and residing together; all of such persons (including related subfamily members) are considered as members of one family.

Non-family Householder. A non-family householder is a person maintaining a household while living alone or with non-relatives only.

B Bootstrap Robustness

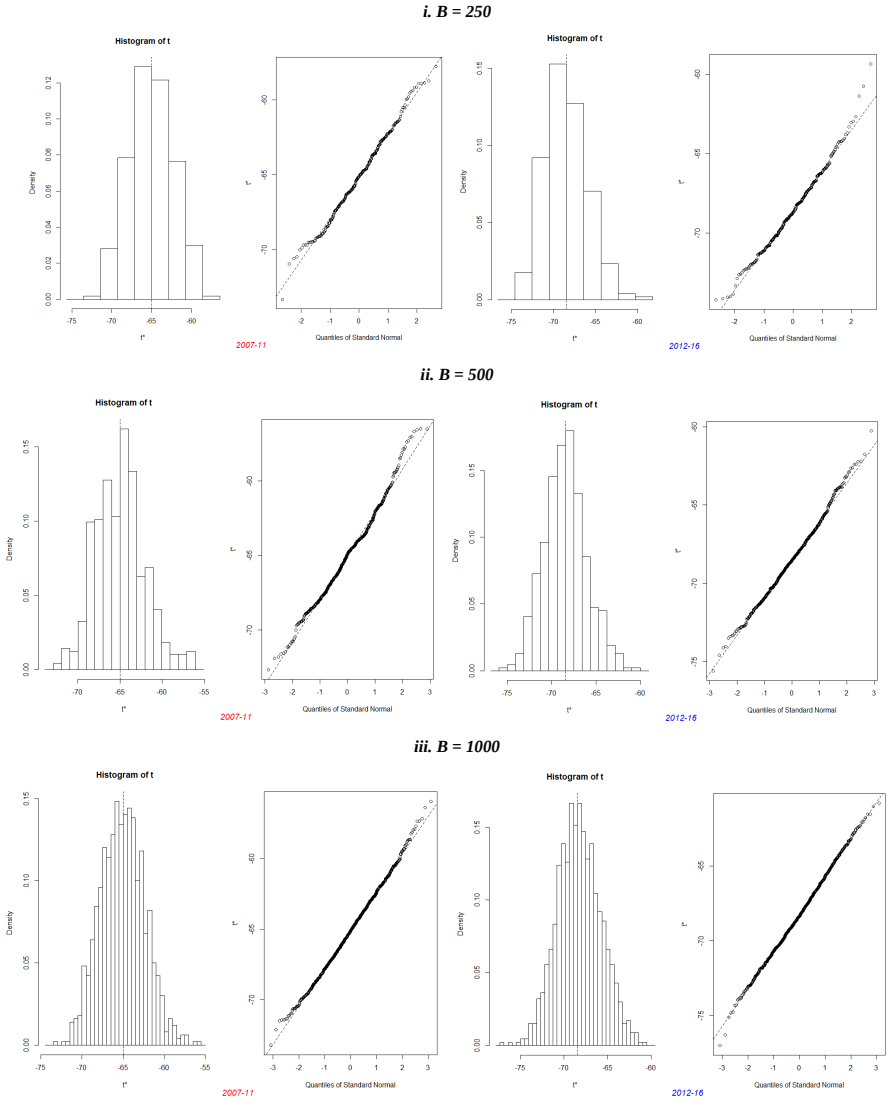


Figure 4: **Bootstrap Robustness for varying number of bootstrap sample**

Note 1: The figure shows the bootstrap distributions of test statistic Δ for number of samples, B = 500, 1000, 5000. In each panel, the left column is the plot for 2007-11 and the right panel is the plot for 2012-16.
Note 2: This check confirms the qualitative robustness of the bootstrap algorithm. The sampling distribution of the test statistic, Δ is approximately normal and a t-test can be used to test for difference of means.

C Census Regions

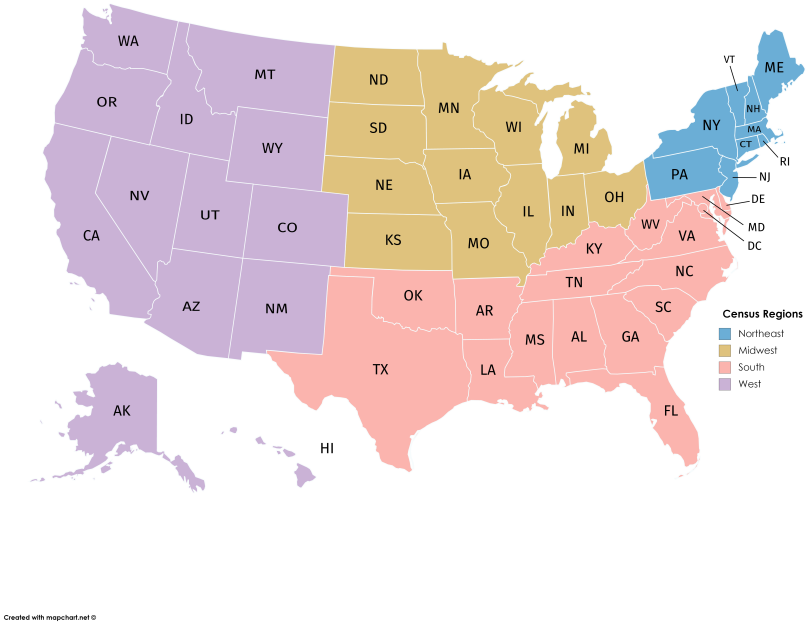


Figure 5: **Four Census Regions**

Source: https://mapchart.net/usa.html
Notes:
Northeast: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont, New Jersey, New York, Pennsylvania.
Midwest: Illinois, Indiana, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota.
South: Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma, Texas.
West: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming, Alaska, California, Hawaii, Oregon, Washington.