Misestimating Home Values: Consequences for Household Finance*

Stefano Corradin† José L. Fillat‡ Carles Vergara-Alert§

Abstract

This study examines the effect of systematic household misestimation of home prices on financial decisions, including stock holdings, consumption, and asset allocation. Using exogenous variation in house values, mortgage debt, and homeowner misestimation identified through differences in local housing market characteristics, we find that a $60,000 increase in house overvaluation (approximately one standard deviation) results in a 1.4-1.9% decrease in risky stockholdings, a 1.5-4.3% increase in consumption, and a 1.3-2.5% increase in the share of risk-free assets over liquid wealth. The results highlight the need to better understand how housing wealth and beliefs about housing values affect portfolio choice, spending, and overall household finance.

JEL Classification: G11, D11, D91, R21, C61.

Keywords: Household Finance, Portfolio Choice, Housing, Misestimation.

*IESE IRB approval ref. IESE.2024.08. We would like to thank Antonio Davila for his research assistance. We thank Fernando Alvarez, Laurent Calvet, Joao Cocco, Pierre Collin-Dufresne, Marjorie Flavin, Tim Landvoigt, Claus Munk, Wayne Passmore, Joe Peek, Federico Perali, Eric Smith, Jacob Sagi, Chester Spatt, Richard Stanton, Suresh Sundaresan, Lars Svensson, Selale Tuzel, Stijn Van Nieuwerburgh, Nancy Wallace, and Scott Wang for their helpful comments. The views expressed in this paper are those of the authors and do not necessarily represent the views of the European Central Bank, Federal Reserve Bank of Boston, or Federal Reserve System. Vergara-Alert acknowledges financial support from the Social Trends Institute Foundation 2022, the State Research Agency of the Spanish Ministry of Science, Innovation and Universities MCIN/AEI/10.13039/501100011033 ERDF (Grant Ref. PGC2018-097335-A-I00), and NextGenerationEU/PRTR (Grant Ref. TED2021-131238B-I00).

†European Central Bank, Sonnemannstrasse 20, 60314 Frankfurt am Main, Germany. E-mail: stefano.corradin@ecb.europa.eu.

‡Federal Reserve Bank of Boston, 600 Atlantic Ave., Boston, 02118 MA. E-mail: jose.fillat@bos.frb.org.

§IESE Business School. Av. Pearson 21, 08034 Barcelona, Spain. Email: cvergara@iese.edu.

1 Introduction

Housing represents the most important asset for most households. As such, house values play a key role in decisions about household choices including stock holdings and consumption. However, households' own estimates of their house values are often not aligned with market prices. Although the average house value misestimation across all home-owning households is close to zero (7,600 dollars of undervaluation on average in our sample), its standard deviation is large (59,800 dollars). We find that 5 percent of the homeowners undervalue their houses by at least 87,500 dollars, while 5 percent overvalue their houses by at least 53,000 dollars.

House value misestimation has been documented for over half a century.[1] However, the literature on portfolio choices with housing assumes that households accurately observe house prices (see for example Flavin and Yamashita (2002), Campbell and Cocco (2003), Cocco (2005), Yao and Zhang (2005), Fischer and Stamos (2013), Corradin, Fillat, and Vergara-Alert (2014), Chetty, Sándor, and Szeidl (2017) and Chen, Michaux, and Roussanov (2020)).

In this paper, we study how house value misestimation affects households' portfolio and consumption decisions. We first develop a simple theoretical framework to show the implications of incorporating misestimation in the analysis of household choices in the presence of housing. We then use household level data to estimate the effects of misestimation on portfolio choices and consumption decisions.

We measure house value misestimation as the difference between the owner's subjective valuation of her house relative to its market value, which is adjusted for home improvements.[2] We exploit a new mechanism based on homeowners who just purchased a house. Our key assumption is that the house's market value is known with certainty only at the time of purchase; that is, misestimation is zero at the time of purchase. After purchase, the market value of the house follows a random process that the homeowner can estimate but does not accurately observe. Using this assumption, we create a novel measure of misestimation by comparing data on self-reported (subjective) housing values from the Panel Study of Income Dynamics (PSID) with market house prices constructed using ZIP code level transaction-based house price indexes from FHFA. Our

[1] Kish and Lansing (1954) and Kain and Quigley (1972) find large discrepancies when they compare homeowners' reported house values to values obtained from professional appraisals. These two studies implicitly assume that appraisals are free of error. Robins and West (1977) also assume that appraisals are unbiased estimates of house values and conclude that house values determined by homeowners and professional appraisals contain errors of 7 percent and 5 percent, respectively. Although there is a consensus about the existence of measurement errors in house prices, there is no agreement on the sign and magnitude. Kish and Lansing (1954), Robins and West (1977), Ihlanfeldt and Martínez-Vázquez (1986), Goodman Jr. and Ittner (1992), Kiel and Zabel (1999), Agarwal (2007), and Benítez-Silva et al. (2015) document the overestimation of reported house values, which range from 3 percent to 16 percent. In contrast, the empirical analyses in Kain and Quigley (1972) and Follain and Malpezzi (1981) find that owners' self-reported house values underestimate house prices by about 2 percent.

[2] Throughout the paper, we use "misestimation" and "house value misestimation" interchangeably. This misestimation is directional: positive misestimation corresponds to overvaluation and negative misestimation corresponds to undervaluation. We do not consider misestimation in any other asset class.

novel measure shows that misestimation is a widespread and sizable phenomenon, with substantial variation across households and regions, but its average is close to zero.

To guide our empirical analysis of the effects of house value misestimation on portfolios and consumption, we develop a stylized three-period model of portfolio choice that incorporates household misestimation of housing values, building on the framework established by Cocco (2005). Our model relates to an emerging literature that incorporates survey evidence on house price expectations (Kuchler, Piazzesi, and Stroebel (2023)) into models with housing as well as previous work that emphasizes the effect of expectations on housing demand (Landvoigt (2017)), mortgage level choices (Bailey et al. (2019)), and home improvements (Choi, Hong, and Scheinkman (2014)).[3] Recent papers examine the role of house price expectations in shaping key housing decisions, including the rent-versus-buy choice Bailey et al. (2018), the timing of home sales Bottan and Perez-Truglia (2025), and characteristics of home purchases such as price and size Gargano, Giacoletti, and Jarnecic (2023).

In our model, households face risky home prices and can invest in risky stocks, in addition to risk-free assets. After buying the house via a mortgage, households hold heterogeneous beliefs about the growth rate of house prices. We assume that households' beliefs about the expected growth rate of house prices follow a normal distribution where the mean is the expected growth rate of market value housing prices. These beliefs determine the level at which households estimate their future home values, and therefore affect their investment and consumption decisions.

This modeling approach allows us to establish a causal relationship between the *exogenous* changes in house value misestimation, defined as the difference between the expected value of house price based on the household's mean belief and the expected value of market value house price based on the market mean, and household's decisions that we employ in the empirical analysis later. The model provides the following predictions. First, households overvaluing their home perceive that their overall risk exposure as already high, and consequently prefer to reduce the stock share of liquid wealth with constant relative risk aversion preferences. Second, households increase non-housing consumption and hold more safe financial assets rather than expose themselves to additional stock market risk.

We then empirically investigate the relationship between households' house value misestimation and their portfolio and consumption choices. We use the PSID household-level data from 1984 to 2021 to find that a $60,000 increase in house overvaluation (approximately one standard deviation) results in a 1.4-1.9% decrease in risky stockholdings, a 1.5-4.3% increase in consumption, and a 1.3-2.5% increase in the share of risk-free assets over liquid wealth, holding house value and mortgage debt constant.

Our empirical approach builds on (Chetty, Sándor, and Szeidl 2017), who underscore the im-

[3] Building on macroeconomic general equilibrium frameworks, Burnside, Eichenbaum, and Rebelo (2016) and Kaplan, Mitman, and Violante (2020) incorporate house price expectations to study aggregate boom-bust dynamics in housing markets.

portance of separately identifying home equity and mortgage debt when analyzing the impact of housing on portfolio decisions. Following their methodology, we employ two distinct instruments: a measure of state-level housing supply elasticity for the subjective market value of the house, and the interaction of state-level average house prices at the year of purchase with national mortgage rates for the outstanding mortgage balance. While our analysis draws on a different dataset and covers a longer time period, we verify that–prior to incorporating house price misestimation–our baseline estimates are consistent with those of Chetty, Sándor, and Szeidl (2017) in both sign and magnitude. This alignment lends credibility to our empirical strategy.

Our empirical analysis extends the existing literature by explicitly accounting for the role of house value misestimation in portfolio decisions. Since misestimation itself may be endogenous, we employ an instrumental variable strategy to identify its effect. Therefore, we introduce two novel instruments that capture variation in households information sets based on locally available signals. The first instrument is the number of housing transactions at the ZIP code level, under the assumption that while local transaction volume shapes information availability, it does not directly influence individual portfolio choices. As an alternative, we use the volume of local Google searches related to housing transactions, which similarly proxies for information exposure. Our results remain robust across both instruments.

Finally, a fundamental assumption in our empirical approach is that the only way to completely mitigate misestimation is to have the house on sale continuously on sale, and receive periodic market offers from buyers. One could argue that the presence of online real estate databases like Zillow, professional appraisals for refinancing or home equity extraction, and municipalities' real estate tax assessments, should mitigate house price misestimation. The main problem is that these estimates of market valuation are not exempt from error and rarely coincide with actual transaction prices. Zillow's website documented, in 2018, that 15.7 percent of the Zillow market estimates miss the subsequent transaction price by more than 20 percent, and 50 percent of the estimates miss the transaction price by more than 5 percent.[4,5]

The paper is structured as follows. In Section 2 develops our measure of misestimation and documents its stylized facts. In Section 3 we describe the stylized model that guides our empirical approach and we study its comparative statics. In Section 4 we develop our empirical strategy. Section 5 presents and discusses our empirical results. Finally, Section 6 concludes.

[4]Source: https://www.zillow.com/zestimate/#acc.

[5]Zillows automated valuation model (the "Zestimate") was introduced only in 2006 and, at launch, covered about 40 million U.S. homes; nationwide coverage expanded only gradually over the subsequent years. Consequently, online real-estate databases such as Zillow were either absent or too thinly populated to affect the properties included in our sample.

2 House Value Misestimation

The main goal of our paper is to analyze the effects of housing value misestimation on stock holdings, consumption, and housing decisions. The empirical analysis in this paper is conducted using several sources of data. First and foremost, we use PSID data from 1984 to 2021 to obtain information at the household level on stock holdings, consumption, and housing decisions. PSID contains a panel of individuals and households that are followed over time. The most relevant variable is the self-reported value of the households' homes. The survey also provides socio-economic characteristics of the households and granular geographic location.[6] Specifically, we use data on family income; family size (number of family members); and the head of household's age, gender, education, marital, and employment status.

We estimate house value misestimation as the difference between the household's subjectively determined house value and the house's actual market value. Self-reported values in PSID are our measure of *subjective house values*. We use the Federal Housing Finance Agency (FHFA) House Price Index (HPI) at the five-digit ZIP code level to construct a proxy for the house's *market value*. We estimated the market value of the property by applying the local HPI growth rate for the household's ZIP code to the most recent purchase price of the house. Formally, misestimation $m_{i,t}$ for each household i at time t is defined as $m_{i,t} = HV_{i,t}^S - HV_{i,t}^M$, where $HV_{i,t}^S$ denotes the subjective value of the house i at time t that the owner reported in the PSID and $HV_{i,t}^M$ is the market value of the house. A positive value of $m_{i,t}$ indicates overvaluation, while a negative value indicates undervaluation. We make one key assumption to build our measure of house value misestimation: the house value that households report on the year of purchase corresponds to the true market value of the house. Hence, household i's house value misestimation is zero at the time of the housing transaction, or $HV_{i,t_0}^M = HV_{i,t_0}^S$. Thereafter, home improvements add to the market value of the house, which evolves following the price growth in the corresponding ZIP code level. Specifically, $HV_{i,t}^M = \left(HV_{i,t-1}^M + HI_{i,t-1}\right) \Delta \log(HPI_{zip^i,t})$, where $HI_{i,t-1}$ denotes the home improvements for household i at time $t-1$. This assumption allows us to use a repeat-sales index at a very granular level (i.e., at the ZIP code level) as opposed to using a hedonic pricing model to account for the house's market price.

Table 1 reports the main descriptive statistics of our measure of misestimation. We observe that, on average, households tend to slightly underestimate the value of their homes, by about $\$7,600$ in our sample, with a standard deviation of $59,800. Remarkably, five percent of the households undervalue their house by more than $\$87,500$, and another five percent overvalue their house by more than $\$53,000$. Figure 1 shows the empirical distribution of misestimation. The distribution appears relatively symmetric, with substantial variation across households, suggesting that both

[6]We use the restricted Geospatial Data Tract Level, produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan. This panel dataset contains the census tract info and ZIP code location of each household. See Appendix OA-I for more details about the data construction.

Table 1: **House Value Misestimation Statistics.** Sample average, standard deviation, percentiles 5 percent and 95 percent, and number of observations for house value misestimation. The data on subjective house values and house improvements comes from PSID. Data on market house values is derived using FHFA data at the five-digit ZIP code level and PSID home prices at transaction times. Period: 1984-2021.

	Mean	Std. Dev.	p5	p95	Obs.
House Value Misestimation ($\times\$100,000$), $m_{i,t}$	-0.076	0.598	-0.875	0.530	60,901
Subjective House Value ($\times\$100,000$), HV^S	1.750	1.796	0.146	5.000	60,901
Market House Value ($\times\$100,000$), HV^M	1.866	2.384	0.148	5.630	60,901
Home Improvements ($\times\$100,000$), HI	0.026	0.200	0.000	0.130	60,901
HI if ($HI > 0$)	0.412	0.688	0.100	1.400	38,651

overestimation and underestimation of home values are common.

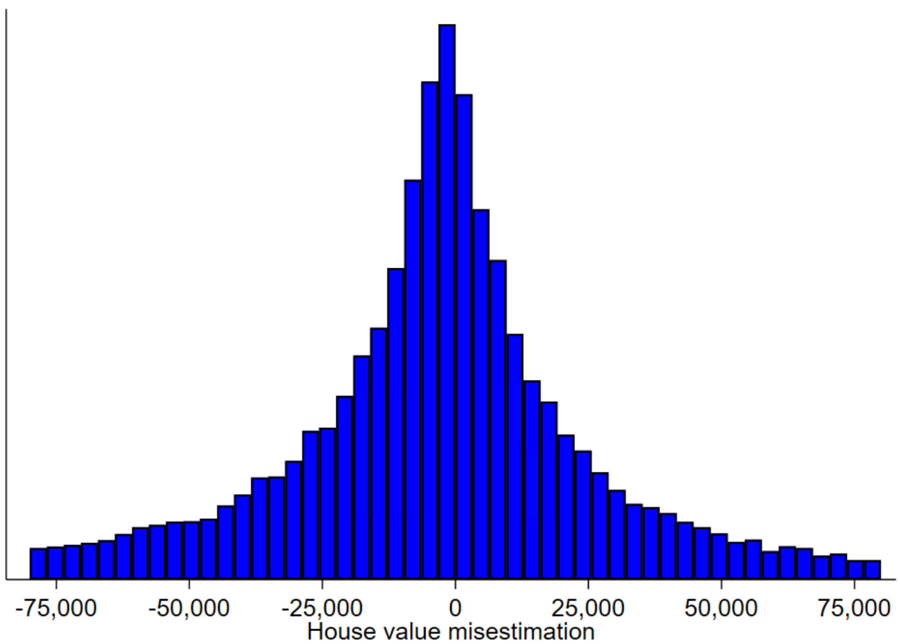

Figure 1: **Distribution Plot of House Value Misestimation.** This histogram shows the empirical distribution of house value misestimation for U.S. households between 1984 and 2021. House value misestimation is calculated as the difference between a household's self-reported house value in the Panel Study of Income Dynamics (PSID) and its estimated market value, which is derived using the Federal Housing Finance Agency (FHFA) House Price Index (HPI) at the five-digit ZIP code level. A positive misestimation value indicates overvaluation by the homeowner, while a negative value indicates undervaluation.

We investigate whether the extent of misestimation is driven by key socioeconomic factors, including family income, household size, age, gender, educational attainment, marital status, and employment status. The results, presented in Table 2, indicate that, although there is some persistence (captured by the positive and significant sign of the coefficient $m_{i,t-1}$), none of these variables

Table 2: **Misestimation and socioeconomic indicators.** The dependent variable for all specifications is misestimation (in $100,000), $m_{i,t}$. We control for the logarithm of family income, number of family members, gender (male=1), education (high school or more=1), marital status (married=1), and employment status (employed=1) of the head of the household. Column (1) also controls for lagged misestimation, $m_{i,t-1}$. All our estimations use age of the head of the household, year and 5-digit ZIP code-level fixed effects. Robust t-statistics are shown in parentheses. ****, **, and * indicate significance at the 1%, 5%, and 10% level of confidence, respectively.

	(1) $m_{i,t}$	(2) $m_{i,t}$
$m_{i,t-1}$	0.423***	
	(0.0592)	
Family Income (log)	-0.0291*	-0.0185
	(0.0157)	(0.0198)
Family Size	-0.0052	-0.0100
	(0.0054)	(0.0081)
Gender	-0.0456	-0.0978**
	(0.0291)	(0.0401)
Education	-0.0174	-0.0213
	(0.0189)	(0.0255)
Married	0.00721	0.0270
	(0.0286)	(0.0365)
Employed	-0.0249	-0.0278
	(0.0176)	(0.0258)
Observations	28,101	28,269
R-squared	0.591	0.381
Zip Code FE	Yes	Yes
Year FE	Yes	Yes
Age FE	Yes	Yes

exhibit statistical significance in explaining variations in the level of misestimation.[7] Moreover, a variance decomposition analysis using socioeconomics variables reveals that, family income (in logs), employment status, education, and family size explain 55.8%, 20.2%, 11.7%, and 9.2% of the explained variation, respectively. Therefore, four variables are needed to explain most of the variation attributed to socioeconomic factors, which together only explain 38.4% of the total variation in misperception.

Overall, a key takeaway from our analysis is that house value misestimation is a widespread phenomenon with substantial variation across households. While self-reported home values tend to be slightly underestimated on average, both overestimation and underestimation are common, with a significant fraction of households exhibiting large misestimation errors. Our findings suggest that socioeconomic characteristics such as family income, employment status, education, and household size do not play a significant role in explaining the extent of misestimation. The persistence of misestimation over time, as indicated by the significance of lagged values, highlights the potential

[7]We report their descriptive statistics in Table A-1 in the Appendix.

for systematic biases in the perceptions of property values at the household level. These results underscore the importance of considering misestimation in economic models of household financial behavior, as it may have far-reaching implications for stock holdings, consumption, and housing decisions.

3 Model

3.1 Set-up

We build on a stylized model of housing and portfolio choice as in Cocco (2005) and Chetty, Sándor, and Szeidl (2017), introducing misestimation in house prices. Our model has three dates $t = 0, 1, 2$. A household endowed with house H_0, mortgage debt M_0, and liquid wealth L_0 makes financial investment decisions at $t = 0$ and $t = 1$, and consumption takes place at $t = 1$ and $t = 2$. The household's utility depends on adjustable consumption C_t and housing consumption H_0. The household faces three sources of uncertainty. First, home prices are risky. Second, after $t = 0$ the household holds heterogeneous beliefs about the growth rate of house prices. Third, the agent can invest in a risky asset.

The household i maximizes lifetime expected utility:

$$\max_{\alpha_0,\alpha_1^i,C_1^i,C_2^i} \delta E_0 \left[\frac{\left(C_1^{i,1-\beta} H_0^{\beta} \right)^{1-\gamma}}{1-\gamma} \right] + \delta^2 E_0 \left[\frac{H_0^{\beta(1-\gamma)}}{(1-\gamma)} \times (\underbrace{W_2 - P_2 \times H_0}_{C_2^i})^{(1-\beta)(1-\gamma)} \right]$$
$$+ \delta^3 E_0 \left[\frac{(P_2 \times H_0 + S_2)^{1-\gamma}}{1-\gamma} \right]. \tag{1}$$

At $t = 0$ the household can invest in a risk-free financial asset with return $1 + R_f = \exp(r_f)$ and a risky asset (stocks) with return $1 + R_s = \exp(r)$, where r is normally distributed with mean μ_r and variance σ_r^2. The only choice variable at $t = 0$ is α_0, the share of liquid wealth invested in the risky asset. Let $R_w = \alpha_0 R_s + (1 - \alpha_0) R_f$ denote the household's financial return on liquid wealth and assume that short-sales constraints restrict $\alpha \in [0, 1]$.

Home prices are $P_0 = 1$ and $P_1 = \exp(p_1)$, where p_1 is normal with mean μ_p^i and variance σ_p^2. The correlation between home price growth and stock returns is $\rho = \text{corr}[p, r]$. As in Bailey et al. (2019), we assume that households beliefs about the expected growth rate of house prices follow a normal distribution with mean μ_p^m, the expected growth rate of market value housing prices, and standard deviation σ_p^m

$$\mu_p^i \sim N(\mu_m, \sigma_m). \tag{2}$$

We adopt the normal distribution for our parametric predictions based on the empirical evidence about the distribution of misestimation that we observe in our Figure 1. This simple set-up allows to measure house value misestimation. At $t = 0$ the homeowner has just bought the house H_0 at the price P_0. Immediately after, the homeowner draws their beliefs about the house appreciation from (2), deviating from the market value of housing prices. When $\mu_p^i > \mu_m$ ($\mu_p^i < \mu_m$) homeowners overestimate (underestimate) house prices leading to positive (negative) house value misestimation.

At $t = 1$ the household's budget constraint is

$$C_1^i + P_1 \times H_0 = (1 + \underbrace{R_w}_{\alpha_1^i R_s + (1-\alpha_1^i) R_f}) \times L_0 + Y_1 + P_1 \times H_0 - (1 + R_{MTG}) \times \frac{M_0}{2}, \tag{3}$$

where R_{MTG} is the mortgage rate, Y_1 is the labor income, which we assume is deterministic, L_0 is the initial liquid wealth. We assume that the agent repays back half of the mortgage balance increasing her home equity share. The share of the risky asset out of liquid wealth α_1^i and numeraire consumption C_1^i crucially depends on whether the homeowner is optimistic or pessimistic about housing return via μ_p^i.

At $t = 2$ we have two components for the agent's expected utility (see Equation 1). The first component depends on the numeraire consumption decision $C_2^i = W_2 - P_2 \times H_0$. For tractability, we assume that the agent's belief μ_p^i realized between $t = 0$ and $t = 1$ does not change between $t = 1$ and $t = 2$. The second component $P_2 \times H_0 + S_2$ depends on the total market value of assets and addresses the concern that the agent cannot monetize the house at the end of $t = 2$ in our framework by introducing a bequest motive. Following Cocco (2005), we assume that the agent bequeaths the house as well as any unconsumed savings, S_2 to its offspring, who derive CRRA utility from the total market value of these assets.

3.2 Parameters

Before presenting the numerical results, we first specify the parameter values for our model. For parameters related to life-cycle portfolio choice, we follow Chetty, Sándor, and Szeidl (2017) setting the risk-free rate at $R_f = 0.02$, the stock risk premium at 0.06, and the annual stock return volatility at $\sigma = 0.157$ per annum. The mortgage rate is set as R_{MTG} at 0.04. We assume that $\gamma = 10$. Regarding housing preferences, we set the relative preference parameter at $\beta = 0.3$. We set the parameter of expected housing return at $\mu_p = 0.016$ as a base parameter. We set the house price volatility σ_p at 0.062 and the belief dispersion σ_m at 0.01. Both Cocco (2005) and Yao and Zhang (2005) assume a zero correlation between housing and the stock market, $\rho = 0$.[8]

We set the time horizon of our model to 10 years to represent an investment horizon over which housing commitments are likely to be important. We set liquid wealth $L_0 = \$44,000$, home value $P_0 \times H_0 = \$150,000$, and mortgage $M_0 = \$105,000$ implying an initial loan-to-value ratio of 70%. Finally, the present value of future labor income is approximately five times current financial wealth

[8]This assumption is also supported by Figure A-1 in the Appendix.

for households in their late 40s and early 50s, and hence we set $Y_1 = 5 \times L_0$ as in Chetty, Sándor, and Szeidl (2017).

3.3 Predictions

We next study how house value misestimation impact investment in the risky asset and numeraire consumption. Figure 2 displays the distribution of house value misestimation based on the model at $t = 1$. We compute misestimation as the difference between the expected value of house price based on μ_p^i and the expected value of market value house price based on μ_m

$$m_1^i = E_1[P_2|\mu_p^i] - E_1[P_2|\mu_m], \tag{4}$$

which is consistent to our definition of empirical house value misestimation presented in Section 2. The first component based on average belief μ_p^i corresponds to the subjective value of the house that the owner reported in the PSID, while the second component based on market expected growth rate μ_m corresponds to the market value of the house. The model distribution closely mimics the empirical distribution of misestimation that we observe in Figure 1. When $\mu_p^i > \mu_m$ ($\mu_p^i < \mu_m$) we have positive (negative) house value misestimation $m_1^i > 0$ ($m_1^i < 0$).

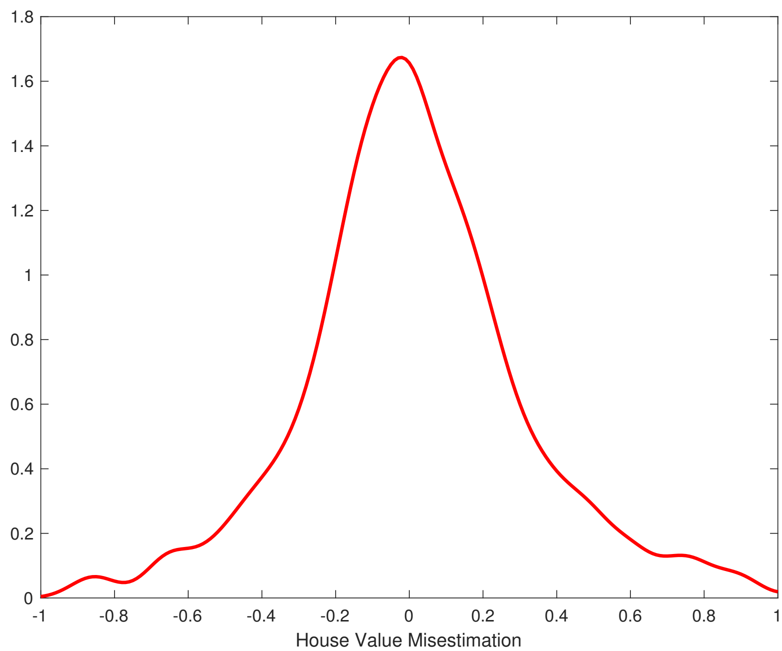

Figure 2: **Distribution Plot of House Value Misestimation Based on the Model.** This figure plots the histogram of the variable house value misestimation computed at $t = 1$. Misestimation m_1^i is computed as the difference between the expected value of house price based on μ_p^i and the expected value of market value house price based on μ_m.

We first focus our analysis on the relationship between misestimation and investment in the risky asset. We solve the model numerically, as closed-form solutions are not available.[9] Figure 3 shows the numerical results household's allocation to the risky asset as a function of house value misestimation. It shows that investment in the risky asset is declining in misestimation m_i. An increase in the average belief μ_p^i increases the perceived net return on housing, leading to positive misestimation m_1^i and induces the agent to reduce the exposure to the stock market, therefore choosing a lower share in the risky asset α_1^i. As a result, the agent increases her exposure to the the risk-free asset, $1-\alpha_1^i$, when misestimation increases, as it is also shown in Figure 3. Prediction 1 summarizes this result:

Prediction 1: *At time* 1 *a household with a higher average belief* μ_p^i *overestimates the value of the house, resulting in a larger misestimation* m_1^i *and a lower allocation to the risky asset* α_1^i.

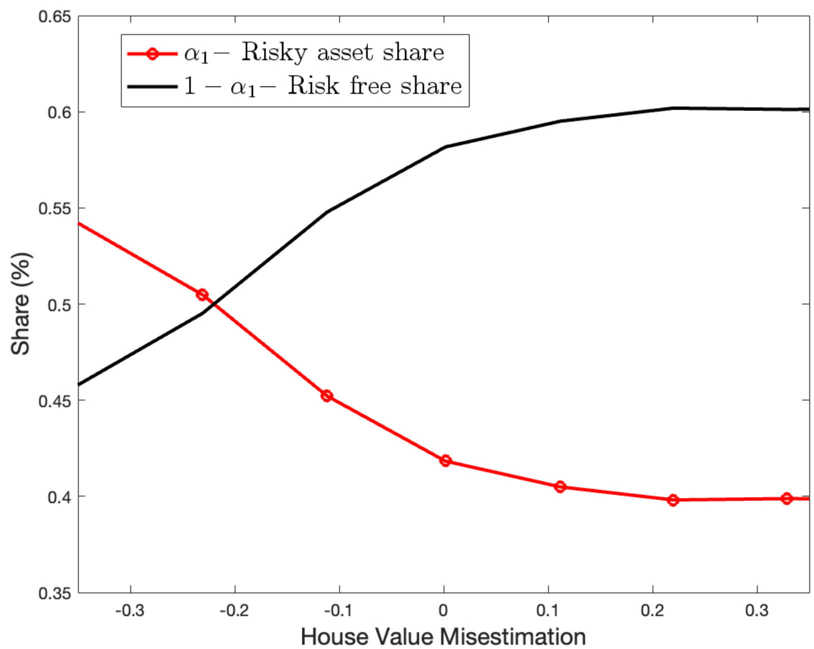

Figure 3: **Relationship Between House Value Misestimation and Risky and Risk-Free Asset Allocation.** This figure plots the household's optimal allocation to risky assets (e.g., stocks) and risk-free assets as a function of house value misestimation, derived from the numerical solution of the model. House value misestimation is defined as the difference between the homeowners subjective valuation and the market value of the house.

Second, we focus on the relationship between misestimation and consumption. Figure 4 shows the numerical results household's consumption as a function of house value misestimation. It

[9] We use the same numerical techniques as Cocco (2005) and Chetty, Sándor, and Szeidl (2017) to solve the model. We use backward induction and compute continuation values over grids. We approximate the state and choice variables using equal-spaced grids, and the probability density functions of shocks with Gaussian quadratures.

shows that consumption C increases in misestimation m_i. As misestimation increases (that is, as households increasingly overvalue their homes), consumption rises. Overestimating homeowners perceive themselves as wealthier and therefore consume more, while underestimating households consume less. An increase in the average μ_p^i increases the perceived net return on housing, leading to positive misestimation m^i, which induces the agent to consume more. Prediction 2 summarizes this result:

Prediction 2: *At time* 1 *a household with a higher average belief* μ_p^i *overestimates the value of the house, resulting in a larger misestimation* m_1^i *and a higher numeraire consumption* C_1^i.

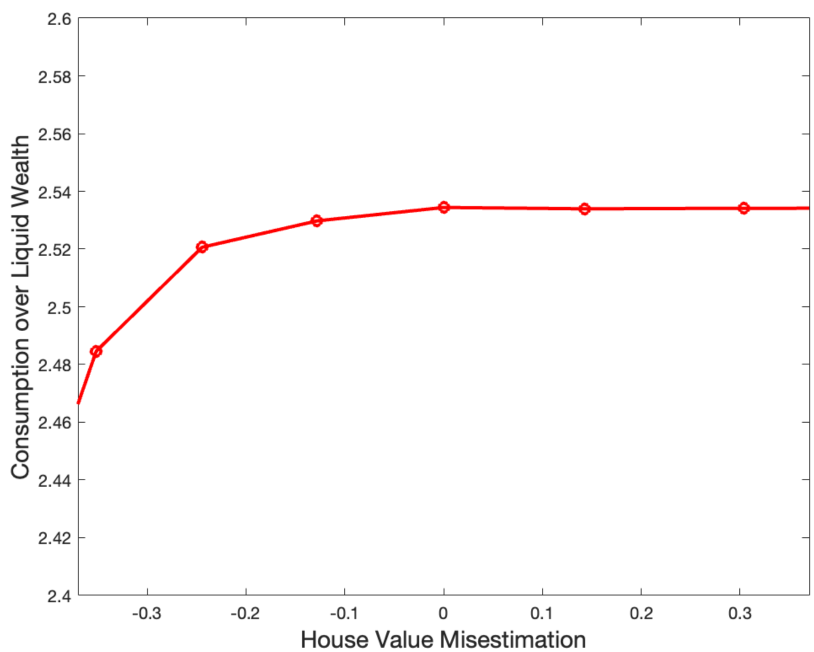

Figure 4: **Relationship Between House Value Misestimation and Consumption.** This figure illustrates the model's predicted relationship between house value misestimation and non-housing consumption, expressed as a share of liquid wealth.

4 Empirical Strategy

Following the theoretical predictions in Section 3, our objective is to empirically establish the causal effect that house value misestimation has on households' choices, such as stock holdings and stock market participation, investment in risk-free assets, and nonhousing consumption decisions. We estimate the effects of misestimation on households' choices using the following linear specification:

$$Y_{it} = \beta_1 m_{it} + \beta_2 HV_{it} + \beta_3 MTG_{it} + \beta_4 X_{it} + \eta_t + \eta_{\text{state}} + \eta_{\text{age}} + \varepsilon_{it}, \quad (5)$$

where Y_{it} is the variable of interest for household i at time t, m_{it} is house value misestimation as described in the previous section. HV_{it} and MGT_{it} denote the reported house value, and the mortgage debt, respectively. As in (Chetty, Sándor, and Szeidl 2017), we separate the reported house value from the mortgage debt because they serve distinct roles in household portfolio decisions and have different implications for risk exposure, liquidity, and wealth accumulation.[10] We also include a vector X_{it} of socio-economic controls at the household-year level, including number of family members, family income (in logs), gender, education, and marital status.

All our specifications include time (η_t), state of residence (η_{state}), and age fixed (η_{age}) effects to control for aggregate common trends and unobserved geographical variation. Specifically, age fixed effects enable us to account for all unobservable characteristics and systematic differences across households of the same age cohort that might influence decision-making. This method ensures that our estimation isolates the variation in decisions attributable to house value misestimation, avoiding biases that might arise from omitted variables correlated with age. Thus, our estimates capture the impact of house value misestimation while controlling for potential confounders tied to cohort-specific behavioral patterns, such as generational attitudes towards risk, financial literacy, or typical life-cycle patterns in asset allocation. Moreover, fixed effects offer the advantage of capturing nonlinearities and cohort-specific dynamics that a simple age control might overlook, leading to more robust and interpretable results. Additionally, by including year fixed effects, we account for time-varying factors that affect all households uniformly, such as macroeconomic trends, policy changes, or market shocks, while state fixed effects control for spatial heterogeneity, such as differences in housing markets, state regulations, or local economic conditions. Together, these fixed effects create a robust framework that isolates the variation in household decisions attributable to house value misestimation, free from confounding influences linked to age, temporal dynamics, and geographic disparities.

Crucially, we must isolate exogenous variation in misestimation to address potential endogeneity concerns, measurement error in our misestimation variable, and reverse causality. We do so by using two novel instruments for our house value misestimation variable: differences across the number of housing transactions at the ZIP code level and local Google Trends data on searches related to housing markets. In the following subsection 4.1 we develop these two instruments. In subsection 4.2 we justify the adoption of two well-established instruments for house values and mortgage debt to address potential endogeneity concerns in the choice of house and mortgage size.

4.1 Instrumenting House Value Misestimation

The amount of housing market-related information available to the households can have an impact on their ability to develop more accurate and reliable assessments of property values in the area.

[10]By analyzing mortgage debt separately, we can isolate the net effect of housing on household wealth and portfolio decisions while accounting for the risks introduced by leverage.

Based on this idea, we introduce two novel instrumental variables designed to isolate exogenous variation in misestimation. Both instruments are of similar nature: We use a variable that is strongly correlated with the accuracy of a household's house value estimation. We interact this variable with the sign of the household's previous misestimation (i.e., whether the house was overvalued or undervalued) to capture the effect of information on the absolute value of misestimation. While the availability of housing market-related information is relevant in determining misestimation, they are arguably exogenous to the households' portfolio and consumption decisions.

First, we use the number of housing *transactions* from CoreLogic data at the ZIP code level. The idea is that, the more liquid local house markets are, the more information households have at hand to infer the value of their house. The key assumption is that the number of local transactions do not affect households' portfolio and consumption decisions. We calculate transaction rates per capita by ZIP code. These rates form the basis of percentile bins (10%, 20%, up to 90%) for each year. The resulting categorical variable, *trans10bins*, stratifies ZIP codes into deciles based on transaction volume, reflecting variations in local housing market dynamics. We interact this categorical variable with the sign of misestimation in the previous period, as more information has the potential effect of reducing absolute value of misestimation. This instrumental variable enables us to capture exogenous variation in housing activity within ZIP code, providing a robust control for market shocks that affect household financial decisions. This instrument also mitigates the measurement error in the market value of the house. A higher volume of local transactions systematically lowers the idiosyncratic component of an individual house's market value. In sum, using the number of housing transactions at the ZIP code level as an IV addresses critical endogeneity concerns by providing an exogenous source of variation in house value misestimation.

Second, we use an alternative instrument for robustness purposes. Consistent with the first approach, we employ the number of housing transaction-related *searches* at the most granular level available from Google Trends. We construct a dictionary of keywords associated with households engaged in home buying or selling activities, like "homes for sale", "mortgage rates", "real estate agent near me", etc.[11] This lexicon serves as the basis for identifying relevant search queries on Google Trends, which we then utilize as an alternative measure of real estate market conditions. To create the instrumental variable, we first assign each ZIP code to a decile based on search intensity observed on Google Trends, stratified by year and geographic region. We then interact this ranked search intensity measure with the sign of the previously observed household misestimation. The approach allows us to exploit both the cross-sectional and temporal variation in real estate search activity as an exogenous source of identification, while also accounting for potential nonlinearities in the relationship between search behavior and households' house value estimation.

[11] In Appendix OA-II, we provide the dictionary that we use to obtain the intensity of search by Designated Market Area (DMA), which we map to ZIP codes. DMAs are geographic regions in the US defined by Nielsen Media Research to represent television and radio markets. These regions serve as reference units for Google Trends, which provides search data that can be segmented by such regions.

4.2 Instrumenting House Value and Mortgage Debt

Building upon Chetty, Sándor, and Szeidl (2017), we employ two separate sets of instrumental variables for house value and mortgage debt. We separate the reported house value from the mortgage debt because they serve distinct roles in household portfolio decisions and have different implications for risk exposure, liquidity, and wealth accumulation.

To instrument the subjective house value, we interact a measure of state-level housing supply elasticity with FHFA national house prices. This approach isolates exogenous variation in house prices due to supply constraints, allowing us to estimate the causal effect of changes in property value on financial portfolios. For robustness, we employ two different measures of elasticity of housing supply. The most widely used measure of elasticity is the one developed in Saiz (2010), derived from land availability and regulation data. Alternatively, we use the measure proposed in Guren et al. (2020), which exploits the fact that local house-price sensitivity to regional prices differs across metropolitan statistical areas (MSAs). They construct the instrument by estimating the historical sensitivity of local house prices to regional housing cycles and interacting the historical sensitivity with current shocks to regional house prices, akin to a Bartik-based instrument. The benefit of that instrument is that it helps predict local house prices by exploiting the fact that house prices in some cities are more sensitive to regional fluctuations than house prices in other cities in the same region.

To instrument mortgage debt, we use the year-of-purchase average house prices in the individual's state interacted with mortgage rates. By comparing individuals who purchased homes during different market conditions, we separate the effects of home value from mortgage debt while controlling for overall wealth changes.

These instruments help control for unobserved variation such as local labor market conditions and selection biases in housing purchase timing. By utilizing these instruments, we disentangle the separate impacts of home value and mortgage debt on households' decisions.

4.3 Additional Empirical Challenges

Our empirical strategy addresses endogeneity concerns that may bias the estimation of the effects of misestimation on households' choices. Variables of interest like investment and consumption decisions, misestimation, home values, and mortgage debt could be subject to measurement error, or present reverse causality.

In particular, our constructed measure of misestimation might be subject to measurement error that can lead to attenuation bias. By design, we utilize the household's reported home value accepting this as the representation of their perceived property worth, including home improvements. Hence, the only source of measurement error in the misestimation variable is our proxy for the market value of the house. The source of measurement error is simply the existence of house-specific characteristics that are not captured in the corresponding households' ZIP code HPI. A systematic relationship between house-specific characteristics and households' portfolio or consumption decisions could further bias our estimates.

Reverse causality also presents a significant challenge. The relationship between house value misestimation and household choices may be bidirectional. For instance, a household experiencing financial distress might adjust its perception of its house value or misreport mortgage debt, leading to simultaneity bias.

Our careful selection of instruments described above addresses both measurement error and simultaneity concerns in several critical ways. By using exogenous variation from housing transactions and search intensity, we address the biases related to these challenges. Specifically, the ZIP code-level housing transaction rates provide an external proxy for market dynamics, which reduces the idiosyncratic noise inherent in individual house valuations. This directly addresses measurement error by anchoring perceived house values to observable and systematic market activity, thereby improving the accuracy of our misestimation variable. Moreover, the Google Trends-based search intensity captures local interest in housing market activity, further reinforcing the robustness of our identification strategy by introducing variation that is orthogonal to household-level financial choices.

Simultaneity is also addressed through these instruments, as they are constructed based on external market activity and information flows that are unlikely to be directly influenced by individual household decisions. For instance, the interaction of transaction volume or search intensity with lagged misestimation ensures that the instruments capture external drivers of misestimation rather than endogenous household behaviors. By isolating these external sources of variation, our approach breaks the feedback loop between household decisions (e.g., portfolio allocation or consumption) and perceived house values, effectively addressing reverse causality.

In sum, our instrumentation strategy not only controls for omitted variables that could otherwise confound the relationship between misestimation and household choices but also ensures that the variation used for identification is exogenous, robust, and interpretable. This dual mitigation of measurement error and simultaneity bias strengthens the credibility of our empirical results and provides a reliable foundation for causal inference.

Finally, a potential concern regarding heterogeneity in the effects of misestimation due to differences in household characteristics such as income, education, or financial literacy is resolved by our analysis previously presented in Table 2. Specifically, we have run a regression of misestimation on a comprehensive set of socioeconomic indicators, including income, family size, gender, education, employment status, and others. The results demonstrate that these indicators are not significantly related to misestimation, suggesting that misestimation is not systematically driven by observable household characteristics. This finding implies that the effects of misestimation are unlikely to vary meaningfully across these dimensions, reducing concerns about unobserved heterogeneity biasing our results. In the online appendix, we show additional robustness tables exploring the heterogeneous effects by interacting misestimation with the socioeconomic controls. The results indicate that the average treatment effect estimated in our analysis provides a reliable and unbiased measure of the impact of misestimation.

5 Results

This section presents the empirical analysis of house value misestimation on household financial decisions, including stock holdings, stock market participation, risk-free assets, and consumption.

The results, summarized in Tables 4 through 6, demonstrate that overestimation of house values significantly influences portfolio allocation, investment behavior, and consumption patterns. Specifically, a $50,000 increase in house overvaluation results (holding house value and mortgage debt constant), on average, in a 0.9 to 1.7% percent decrease in the share of risky stock holdings, a 1.3–3.3% percentage point reduction in stock market participation, a 2.3–4.4% percent increase in the share of consumption, and a 1.7–2.2% increase in the share of risk-free asset holdings over liquid wealth. These findings highlight the economic significance of house value misestimation in shaping household financial decisions.

5.1 Household Finance Data

The objective of the paper is to establish a theoretical and empirical relationship between misestimation and household finance decisions on stock holdings, consumption, and risk free assets. We use the same PSID sample from 1985 to 2021 to obtain these variables at the household level over time, and Table 3 shows the relevant descriptive statistics. The measure of stock holdings includes direct stock ownership, IRAs, and annuity holdings. To compute households' liquid wealth, we calculate the risk-free assets at the household level. Risk-free assets comprise bonds, insurance (both net of debt), checking and savings balances, minus the outstanding mortgage principal on the primary residence. We calculate consumption as the sum of the food used at home, food used away from home, and food delivered at home. In line with the literature, we use household wealth to normalize portfolio choices. We calculate liquid wealth as the sum of the household's primary residence value, its second home value (net of debt), business value (net of debt), bonds and insurance assets (net of debt), stock holdings (net of debt), checking and savings balances, IRAs and annuities, less the mortgage principal on the primary residence.

The descriptive statistics of key variables presented in Table A-1 provide an overview of the main financial variables included in our analysis. The dataset includes variables such as total and liquid wealth, mortgage amounts, stock holdings, consumption patterns, and risk-free asset allocations. The sample is restricted to households with liquid wealth exceeding $2,963.3, corresponding to the average monthly salary in the U.S. in 2000. The table reports the mean, standard deviation, 5th and 95th percentiles, and the total number of observations for each variable. It shows substantial variation in household wealth and financial behavior, highlighting the heterogeneity in stock market participation, consumption decisions, and portfolio allocation. For comparability across different survey waves, we exclusively focus on first mortgages.

Additionally, the Online Appendix OA-I contains a more detailed description of the data sources and data cleansing process. Table A-1 in this Appendix presents key socioeconomic indicators such as family income, household size, age, gender, education, marital status, and employment status, which serve as control variables in our empirical analysis.

Table 3: **Descriptive Statistics of Key Variables.** This table presents summary statistics for the primary variables in the analysis, including the sample mean, standard deviation, 5th and 95th percentiles, and the total number of observations. Housing market data are sourced from the Federal Housing Finance Agency (FHFA), while data on house value misestimation, household socioeconomic characteristics, and financial decisions are obtained from the Panel Study of Income Dynamics (PSID). The sample is restricted to households with liquid wealth exceeding \$2963.3 ($LW > 2963.3$), which corresponds to the average monthly salary in the U.S. in the year 2000.

	Mean	Std. Dev.	p5	p95	Obs.
Wealth:					
Total Wealth ($\times\$100,000$), TW	4.083	15.077	-0.085	15.100	36,226
Liquid Wealth ($\times\$100,000$), LW	0.284	11.570	-0.480	4.200	36,447
Household Choices:					
Mortgage ($\times\$100,000$), MTG	0.998	1.078	0.000	3.130	44,089
Stock Holdings over Liquid Wealth, SV/LW	0.254	0.365	0.000	0.980	17,857
Consumption over Liquid Wealth, C/LW	0.487	0.781	0.009	1.999	17,782
Risk-free Assets over Liquid Wealth, RFA/LW	0.876	0.737	0.006	1.818	18,889
Stock Participation, $SV > 0$	0.447	0.497	0.000	1.000	18,889
Participants' Stock Holdings, SV/LW	0.613	0.318	0.050	1.000	7,409

5.2 House Value Misestimation and Stock Holdings

Table 4 examines the relationship between house value misestimation and the proportion of household portfolios allocated to stocks. The results show a negative and statistically significant coefficient for the misestimation variable across all specifications. Households that overvalue their homes allocate a smaller share of their liquid wealth to stocks, consistent with Prediction 1 of our theoretical model.

Our estimates show that a one-standard-deviation increases in house overvaluation (which represents an overvaluation of \$59,800) results, on average, in a 1.14 to 1.86% percent decrease in the share of risky stock holdings, holding house value and mortgage debt constant.[12]

This effect remains robust to controlling for house value, mortgage debt, and socioeconomic factors. Consistent with the model presented in Section 3, households that overvalue their house tend to hold a lower share of risky stocks due a substitution effect. Housing is a large risky asset and, a larger valuation overweights risky asset in the households' portfolio. The optimal response is to reduce their exposure to risky stock holdings, as the house is indivisible. The impact of mortgage values are also noteworthy and in line with the model findings. Households that are more levered (larger mortgages) also reduce their risky stock holdings, as leverage reduces their risk taking capacity. Home equity has an impact on stock holdings comparable to that of overvaluation. (Chetty, Sándor, and Szeidl 2017) emphasize the need to control for leverage when evaluating the effects of home values on households' decisions. Consistent with their finding, we also find that higher home equity also results in a crowding out of risky stock holdings, à la (Cocco 2005). While market house values have a negligible effect or negative effect, mortgage value has a very significant and negative impact on stock holdings. In robustness analysis we find equivalent results with home equity instead of mortgage value. When controlling for home equity, the impact of higher market house value is not significant or negative when using the instruments based on (Guren et al. 2020), exactly in line with (Chetty, Sándor, and Szeidl 2017).

The magnitude of the coefficient estimates on the instrumented misperception are larger than the OLS estimates and remain significant. The Kleibergen-Paap Wald F-statistic confirms the strength of our instrumental variables, lending credibility to the causal interpretation of the estimates. We report the first stage estimates in the Appendix. It is important to reiterate that the choice of instrument enables us to partly address endogeneity and measurement error concerns. Specifications in columns (4) and (7) instrument misestimation with the number of completed housing transactions at the zipcode level. Arguably, households living in a zipcode where more transactions are executed may have more information about the actual market value of the properties, and therefore have a better estimate of the market value of housing in their zipcode (which is our measure HV^S). Similarly, specifications in columns (5) and (8) use as an instrument for misvaluation the state-level number of Google searches containing words related to housing markets. The idea is that more frequent searches result in more informed responses about the household's home values. Both the number of transactions and the search intensity are plausibly uncorrelated with household-level portfolio allocation decisions.

The effects of other variables on stock holdings are also in line with economic intuition. Not all the estimates on the socioeconomic variables are statistically significant, but family income and education have an expected positive and significant impact on stock holdings, as expected.

We explore the extensive margin of stockholdings in Table 5. Our left-hand side of the regression specification is a dummy variable that takes the value of one if the household reports a positive

[12]Note that, from Table 3, one standard deviation in misestimation corresponds to \$59,800.

value of stock holdings in a given year. We keep things simple and run a linear probability model but results are robust to nonlinear specifications like probit or logit.

Our findings indicate that households with overvalued home perceptions are less likely to participate in the stock market. Specifically, we estimate that a one-standard-deviation increase in house overvaluation results (holding house value and mortgage debt constant), on average, in a 1.5–3.59% percentage point reduction in stock market participation. This result is both statistically and economically significant. The OLS results hold and are stronger when we instrument misestimation, market house values, and mortgages. These results on the extensive margin of stock holdings do not have a direct mapping to the model. However the results are consistent with the intuition that overvaluation leads households to reduce their allocation into risky stocks, and with some of the households being on the margin, misestimation pushes them to leave the stock market altogether.

5.3 House Value Misestimation and Consumption

Previous research has underscored limitations in PSID consumption measures (Li et al. 2007; Attanasio and Pistaferri 2014). These challenges remain central in interpreting consumption dynamics from PSID data. Nonetheless, we use the more reliable food-consumption measure (Hall and Mishkin 1982) to further exploit the predictions of household choices in our theoretical framework.

The results in Table 6 suggest that overvaluation has a positive wealth effect on consumption, a result that holds both in the OLS framework of column (2), but also when misvaluation is instrumented to extract the exogenous variation, in columns (4), (5), (7), and (8). We find that an increase of one standard deviation in overvaluation ($59,800) results in 2.63 to 4.31 percentage points higher consumption relative to liquid wealth.

Our results are consistent with the literature quantifying the wealth effect of housing values. The literature has produced several empirical estimates of the marginal propensity to consume (MPC) out of housing wealth. (Poterba 1984) in the found an MPC ranging from 0.04 to 0.06, suggesting a $1 increase in housing wealth would lead to a $0.04 to $0.06 increase in consumption. (Case, Quigley, and Shiller 2005) estimated the MPC to be around 0.05 to 0.15, with higher estimates for the U.S. compared to other countries. (Carroll, Otsuka, and Slacalek 2011) estimated the MPC to be around 0.05 to 0.08, with higher estimates during housing booms. (Mian and Sufi 2011) found that the housing wealth effect was amplified during the 2007-2009 financial crisis, with (Dynan 2012) showing an MPC of around 0.05 to 0.10 during the crisis period. Our primary focus, though, is on the effects of misestimation of house values, and it is not straightforward to compare marginal propensities to consume found in the literature with our estimates. However, for the median household in our sample, whose consumption is 19 percent of their liquid wealth, a $100,000 increase in overvaluation represents 7.2 percentage point higher consumption ratio, that is, an increase from 18 to 26 percent of annual consumption over liquid wealth. For a median liquid wealth of $40,000, consumption increases by $2,880 over a year.

5.4 House Value Misestimation and Risk-Free Assets

Finally, Table 7 analyzes the impact of house value misestimation on the share of risk-free assets over liquid wealth. A $50,000 increase in house overvaluation results (holding house value and mortgage debt constant), on average, in a 1.7 to 2.2% increase in the share of risk-free asset holdings over liquid wealth.

Table 4: **Stock Holdings and Misestimation.** This table shows the effects that house price misestimation (in $100,000) has on a household's stock holdings. The dependent variable is the share of stock value as part of total liquid wealth if the Liquid Wealth is grater than 2963.3. Columns (1) and (2) show the simple OLS results of the panel regressions without and with misperception. Columns (3)–(5) us the (Saiz 2010) elasticity of supply interacted with house prices at the state level as an instrument for the subjective house values, while columns (6)–(8) use the (Guren et al. 2020) measure of elasticity of supply as an instrument for the subjective house values. Misperception is instrumented with a measure of zipcode transactions in columns (4) and (7), and with the intensity of internet searches about housing markets in columns (5) and (8). Mortgage value is instrumented by the HPI index at the year of purchase interacted with long-term interest rates in all specifications. KPW denotes the Kleibergen-Paap Wald F-statistic coefficient, used to assess the strength of instrumental variables in the regression analysis. The $t-$ statistics are reported in parentheses. The symbols ***, **, and * denote the statistical significance of the coefficients at the 99, 95, and 90 percent levels of confidence. Standard errors are clustered at the year and ZIP code level.

VARIABLES	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) IV	(6) IV	(7) IV	(8) IV
$m_{i,t}$		-0.019***		-0.031***	-0.021***		-0.022**	-0.021**
		(0.003)		(0.011)	(0.008)		(0.009)	(0.009)
HV^S	0.031***	0.032***	0.037*	0.054*	0.013	0.029**	0.006	0.002
	(0.001)	(0.001)	(0.019)	(0.031)	(0.020)	(0.013)	(0.012)	(0.013)
MTG	-0.025***	-0.025***	-0.066***	-0.056***	-0.053***	-0.082***	-0.070***	-0.079***
	(0.002)	(0.002)	(0.015)	(0.018)	(0.017)	(0.016)	(0.019)	(0.020)
Family Size	-0.016***	-0.015***	-0.013***	-0.016***	-0.012**	-0.016***	-0.017***	-0.013***
	(0.002)	(0.002)	(0.004)	(0.005)	(0.005)	(0.003)	(0.004)	(0.004)
Family Income (log)	0.039***	0.036***	0.062***	0.026	0.088***	0.075***	0.101***	0.113***
	(0.004)	(0.004)	(0.024)	(0.043)	(0.027)	(0.019)	(0.021)	(0.022)
Gender	-0.017	-0.019*	-0.013	-0.015	-0.018	-0.010	-0.022	-0.009
	(0.011)	(0.011)	(0.014)	(0.018)	(0.018)	(0.015)	(0.019)	(0.019)
Education	0.075***	0.072***	0.086***	0.070***	0.090***	0.074***	0.082***	0.080***
	(0.005)	(0.005)	(0.012)	(0.020)	(0.015)	(0.010)	(0.011)	(0.012)
Married	0.007	0.007	-0.003	-0.016	-0.014	0.002	0.015	-0.008
	(0.010)	(0.010)	(0.012)	(0.015)	(0.015)	(0.014)	(0.017)	(0.017)
Observations	13,724	13,724	8,823	6,954	6,630	7,210	5,712	5,553
R-squared	0.151	0.155	0.073	0.079	0.064	0.041	0.022	-0.016
IV m_i_t	-	-	-	Transactions	Searches	-	Transactions	Searches
IV HV	-	-	Saiz	Saiz	Saiz	γ	γ	γ
IV MTG	-	-	Saiz	Saiz	Saiz	γ	γ	γ
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AGE FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kleibeergen Paap			66.27	34.91	80	91.19	111.6	97.42
KPW			33.15	2.903	6.813	48.57	9.388	8.483
Cragg Donald			29.25	2.647	6.078	74.46	12.84	11.70
Anderson Rubin			9.820	2.285	2.607	13.88	2.994	3.405
Fuller Test			1	1	1	1	1	1
Fisrt Stage F			92.78	63.16	58.83	49.15	34.76	29.22
Hansen J			0	5.899	8.468	0	6.768	7.245

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: **Stock Market Participation and Misestimation.** This table shows the effects that house price misestimation (in $100,000) has on a household's stock market participation. The dependent variable is the stock market participation. Columns (1) and (2) show the simple OLS results of the panel regressions without and with misperception. Columns (3)–(5) us the (Saiz 2010) elasticity of supply interacted with house prices at the state level as an instrument for the subjective house values, while columns (6)–(8) use the (Guren et al. 2020) measure of elasticity of supply as an instrument for the subjective house values. Misperception is instrumented with a measure of zipcode transactions in columns (4) and (7), and with the intensity of internet searches about housing markets in columns (5) and (8). Mortgage value is instrumented by the HPI index at the year of purchase interacted with long-term interest rates in all specifications. KPW denotes the Kleibergen-Paap Wald F-statistic coefficient, used to assess the strength of instrumental variables in the regression analysis. The $t-$ statistics are reported in parentheses. The symbols ***, **, and * denote the statistical significance of the coefficients at the 99, 95, and 90 percent levels of confidence. Standard errors are clustered at the year and ZIP code level.

VARIABLES	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) IV	(6) IV	(7) IV	(8) IV
$m_{i,t}$		-0.025***		-0.060***	-0.038***		-0.036**	-0.033**
		(0.004)		(0.017)	(0.012)		(0.014)	(0.014)
HV^S	0.044***	0.045***	0.081**	0.115**	0.018	0.041**	0.009	-0.009
	(0.002)	(0.002)	(0.031)	(0.054)	(0.030)	(0.019)	(0.018)	(0.019)
MTG	-0.028***	-0.028***	-0.114***	-0.098***	-0.063**	-0.116***	-0.104***	-0.095***
	(0.004)	(0.004)	(0.024)	(0.028)	(0.026)	(0.023)	(0.028)	(0.028)
Family Size	-0.031***	-0.030***	-0.031***	-0.035***	-0.023***	-0.033***	-0.033***	-0.029***
	(0.004)	(0.004)	(0.006)	(0.008)	(0.007)	(0.005)	(0.006)	(0.007)
Family Income (log)	0.075***	0.071***	0.082**	0.004	0.132***	0.126***	0.159***	0.185***
	(0.006)	(0.006)	(0.037)	(0.071)	(0.041)	(0.028)	(0.032)	(0.033)
Gender	-0.024	-0.026	-0.017	-0.017	-0.006	-0.010	-0.025	-0.005
	(0.017)	(0.017)	(0.021)	(0.027)	(0.027)	(0.024)	(0.028)	(0.029)
Education	0.142***	0.139***	0.143***	0.105***	0.167***	0.150***	0.164***	0.166***
	(0.008)	(0.008)	(0.021)	(0.035)	(0.024)	(0.016)	(0.017)	(0.019)
Married	0.022	0.021	0.013	-0.005	-0.020	0.011	0.023	-0.012
	(0.015)	(0.015)	(0.019)	(0.023)	(0.023)	(0.021)	(0.025)	(0.026)
Observations	15,505	15,505	10,066	7,964	7,384	8,222	6,561	6,191
R-squared	0.169	0.171	0.067	0.035	0.082	0.050	0.027	-0.002
IV m_i_t	-	-	-	Transactions	Searches	-	Transactions	Searches
IV HV	-	-	Saiz	Saiz	Saiz	γ	γ	γ
IV MTG	-	-	Saiz	Saiz	Saiz	γ	γ	γ
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AGE FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kleibeergen Paap			73.56	40.32	88.79	106.7	134	121.3
KPW			36.77	3.352	7.525	56.11	11.31	10.67
Cragg Donald			32.91	3.058	6.505	90.75	15.19	14.20
Anderson Rubin			11.90	3.332	2.536	13.88	3.507	3.259
Fuller Test			1	1	1	1	1	1
Fisrt Stage F			134.1	83.06	86.76	78.36	52.89	45.48
Hansen J			0	8.581	7.135	0	9.959	8.837

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: **Consumption and Misestimation.** This table shows the effects that house price misestimation has on a household's allocation to non-housing consumption. The dependent variable is food consumption (Cons) over Liquid Wealth (LW) if the Liquid Wealth is grater than 2963.3. Columns (1) and (2) show the simple OLS results of the panel regressions without and with misperception. Columns (3)–(5) us the (Saiz 2010) elasticity of supply interacted with house prices at the state level as an instrument for the subjective house values, while columns (6)–(8) use the (Guren et al. 2020) measure of elasticity of supply as an instrument for the subjective house values. Misperception is instrumented with a measure of zipcode transactions in columns (4) and (7), and with the intensity of internet searches about housing markets in columns (5) and (8). Mortgage value is instrumented by the HPI index at the year of purchase interacted with long-term interest rates in all specifications. KPW denotes the Kleibergen-Paap Wald F-statistic coefficient, used to assess the strength of instrumental variables in the regression analysis. The $t-$ statistics are reported in parentheses. The symbols ***, **, and * denote the statistical significance of the coefficients at the 99, 95, and 90 percent levels of confidence. Standard errors are clustered at the year and ZIP code level.

VARIABLES	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) IV	(6) IV	(7) IV	(8) IV
$m_{i,t}$		0.025***		0.072**	0.044**		0.041*	0.033
		(0.006)		(0.032)	(0.020)		(0.024)	(0.024)
HV^S	-0.054***	-0.055***	-0.068	-0.152	-0.057	-0.000	0.028	0.035
	(0.004)	(0.004)	(0.058)	(0.115)	(0.051)	(0.038)	(0.036)	(0.038)
MTG	0.041***	0.041***	0.162***	0.157***	0.162***	0.094**	0.080	0.100**
	(0.006)	(0.006)	(0.039)	(0.053)	(0.042)	(0.040)	(0.049)	(0.048)
Family Size	0.100***	0.100***	0.094***	0.106***	0.098***	0.091***	0.095***	0.101***
	(0.006)	(0.006)	(0.013)	(0.018)	(0.015)	(0.013)	(0.014)	(0.016)
Family Income (log)	-0.083***	-0.079***	-0.156**	-0.006	-0.152**	-0.183***	-0.215***	-0.242***
	(0.011)	(0.011)	(0.067)	(0.147)	(0.065)	(0.055)	(0.064)	(0.065)
Gender	0.110***	0.112***	0.148***	0.160***	0.128***	0.069**	0.046	0.033
	(0.028)	(0.028)	(0.032)	(0.038)	(0.034)	(0.033)	(0.035)	(0.038)
Education	-0.120***	-0.117***	-0.110***	-0.030	-0.119***	-0.129***	-0.138***	-0.155***
	(0.013)	(0.013)	(0.037)	(0.073)	(0.038)	(0.029)	(0.032)	(0.034)
Married	-0.019	-0.018	-0.018	-0.015	0.016	0.023	0.021	0.029
	(0.025)	(0.025)	(0.030)	(0.034)	(0.033)	(0.033)	(0.036)	(0.039)
Observations	14,992	14,992	9,693	7,645	7,345	7,930	6,318	6,157
R-squared	0.111	0.112	0.032	-0.008	0.034	0.020	0.004	-0.011
IV m_i_t	-	-	-	Transactions	Searches	-	Transactions	Searches
IV HV	-	-	Saiz	Saiz	Saiz	γ	γ	γ
IV MTG	-	-	Saiz	Saiz	Saiz	γ	γ	γ
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AGE FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kleibeergen Paap			67.69	36.02	87.38	102.2	129.3	120
KPW			33.87	2.996	7.410	53.93	10.97	10.58
Cragg Donald			29.78	2.695	6.425	86.37	14.53	14.05
Anderson Rubin			12.09	3.213	3.127	2.862	2.552	1.715
Fuller Test			1	1	1	1	1	1
Fisrt Stage F			51.08	36.91	39.23	35.56	25.27	24.90
Hansen J			0	13.44	5.684	0	19.12	7.960

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: **Risk Free Assets and Misestimation.** This table shows the effects that house price misestimation has on a household's allocation to risk free assets. The dependent variable is resk free assets (RFA) over Liquid Wealth (LW) if the Liquid Wealth is grater than 2963.3. Columns (1) and (2) show the simple OLS results of the panel regressions without and with misperception. Columns (3)–(5) us the (Saiz 2010) elasticity of supply interacted with house prices at the state level as an instrument for the subjective house values, while columns (6)–(8) use the (Guren et al. 2020) measure of elasticity of supply as an instrument for the subjective house values. Misperception is instrumented with a measure of zipcode transactions in columns (4) and (7), and with the intensity of internet searches about housing markets in columns (5) and (8). Mortgage value is instrumented by the HPI index at the year of purchase interacted with long-term interest rates in all specifications. KPW denotes the Kleibergen-Paap Wald F-statistic coefficient, used to assess the strength of instrumental variables in the regression analysis. The $t-$ statistics are reported in parentheses. The symbols ***, **, and * denote the statistical significance of the coefficients at the 99, 95, and 90 percent levels of confidence. Standard errors are clustered at the year and ZIP code level.

VARIABLES	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) IV	(6) IV	(7) IV	(8) IV
$m_{i,t}$		0.030***		0.042***	0.025**		0.022*	0.022*
		(0.004)		(0.015)	(0.012)		(0.013)	(0.012)
HV^S	-0.047***	-0.048***	-0.035	-0.043	0.015	-0.021	-0.012	0.001
	(0.002)	(0.002)	(0.029)	(0.044)	(0.030)	(0.017)	(0.016)	(0.017)
MTG	0.038***	0.038***	0.071***	0.051**	0.032	0.084***	0.065***	0.060**
	(0.003)	(0.003)	(0.022)	(0.025)	(0.025)	(0.021)	(0.025)	(0.025)
Family Size	0.024***	0.024***	0.016***	0.019**	0.010	0.031***	0.037***	0.032***
	(0.004)	(0.004)	(0.006)	(0.008)	(0.007)	(0.005)	(0.006)	(0.006)
Family Income (log)	-0.032***	-0.027***	-0.074**	-0.047	-0.122***	-0.088***	-0.092***	-0.108***
	(0.006)	(0.006)	(0.034)	(0.058)	(0.039)	(0.026)	(0.027)	(0.029)
Gender	0.003	0.006	0.002	0.007	-0.006	0.010	0.014	-0.006
	(0.016)	(0.016)	(0.019)	(0.024)	(0.025)	(0.022)	(0.026)	(0.027)
Education	-0.104***	-0.100***	-0.127***	-0.108***	-0.147***	-0.119***	-0.119***	-0.121***
	(0.007)	(0.007)	(0.019)	(0.030)	(0.023)	(0.015)	(0.015)	(0.017)
Married	0.012	0.013	0.034**	0.046**	0.055***	-0.004	-0.019	0.004
	(0.014)	(0.014)	(0.017)	(0.020)	(0.021)	(0.019)	(0.022)	(0.023)
Observations	14,861	14,861	9,656	7,659	7,081	7,876	6,307	5,929
R-squared	0.169	0.173	0.071	0.089	0.020	0.037	0.043	0.020
IV m_i_t	-	-	-	Transactions	Searches	-	Transactions	Searches
IV HV	-	-	Saiz	Saiz	Saiz	γ	γ	γ
IV MTG	-	-	Saiz	Saiz	Saiz	γ	γ	γ
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AGE FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kleibeergen Paap			67.93	38.10	84.70	99.53	124	109.7
KPW			33.92	3.159	7.169	52.64	10.52	9.657
Cragg Donald			30.37	2.863	6.230	86.83	14.33	13.18
Anderson Rubin			5.797	2.007	2.130	7.827	1.519	1.895
Fuller Test			1	1	1	1	1	1
Fisrt Stage F			80.38	59.30	51.27	47.40	37.12	31.98
Hansen J			0	5.706	9.653	0	3.704	9.277

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6 Conclusions

This paper provides novel insights into the economic consequences of house value misestimation on household finance. Using a combination of theoretical modeling and empirical analysis, we establish that households systematically overestimate or underestimate the value of their homes, leading to significant shifts in portfolio allocation, consumption, and investment decisions.

We document a robust negative relationship between house overvaluation and the share of risky assets in household portfolios. Specifically, a one-standard-deviation increases in house overvaluation (which represents an overvaluation of $59,800) results, on average, in a decrease of 1.14 to 1.86% in the allocation to risky stock holdings, consistent with our theoretical model's predictions. This finding highlights a critical departure from standard portfolio choice models, which typically assume accurate perceptions of wealth.

Furthermore, we find that overvaluation is associated with increased non-housing consumption. An increase of one standard deviation in overvaluation ($59,800) results in 2.63 to 4.31 percentage points higher consumption relative to liquid wealth. This result underscores the role of housing wealth misperception in the marginal propensity to consume, suggesting that households adjust their spending behavior in response to perceived (rather than actual) wealth gains. Additionally, households with higher perceived home values tend to reallocate financial assets away from stocks towards risk-free assets, reinforcing a conservative shift in portfolio composition.

From an identification perspective, our use of housing market liquidity and real estate search intensity as instrumental variables mitigates concerns of reverse causality and measurement error. The strength of our instruments, confirmed by statistical tests, bolsters the causal interpretation of our results.

These results have broad implications for financial theory and policy. First, they challenge the common assumption in portfolio choice models that households accurately observe their wealth. Second, they suggest that financial advisors and policymakers should account for biases in housing wealth perceptions when designing investment and retirement strategies. Third, given the widespread use of home equity as collateral, our findings imply that misperceptions about home values could have significant implications for credit availability and macroeconomic stability.

Future research should explore the heterogeneity of these effects across demographic groups and different housing market conditions. Additionally, investigating how financial literacy or real-time market information could mitigate the effects of misestimation on household decision-making presents an important avenue for further study. Our analysis suggests that policies aimed at improving the accuracy of homeowners price expectationssuch as more frequent professional appraisals or enhanced financial educationmay help optimize household financial decisions and contribute to overall economic stability.

References

Agarwal, Sumit. 2007. "The Impact Of Homeowners' Housing Wealth Misestimation on Consumption and Saving Decisions." *Real Estate Economics* 35 (2): 135–154.

Attanasio, Orazio, and Luigi Pistaferri. 2014. "Consumption Inequality over the Last Half Century: Some Evidence Using the New PSID Consumption Measure." *American Economic Review* 104 (5): 122–126 (May).

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2018. "The economic effects of social networks: Evidence from the housing market." *Journal of Political Economy* 126 (6): 2224–2276.

Bailey, Michael, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel. 2019. "House price beliefs and mortgage leverage choice." *The Review of Economic Studies* 86 (6): 2403–2452.

Benítez-Silva, Hugo, Selçuk Eren, Frank Heiland, and Sergi Jiménez-Martín. 2015. "How Well Do Individuals Predict the Selling Prices of their Homes?" *Journal of Housing Economics* 29 (C): 12–25.

Bottan, Nicolas, and Ricardo Perez-Truglia. 2025. "Betting on the house: Subjective expectations and market choices." *American Economic Journal: Applied Economics* 17 (1): 459–500.

Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo. 2016. "Understanding booms and busts in housing markets." *Journal of Political Economy* 124 (4): 1088–1147.

Campbell, John Y., and Joao F. Cocco. 2003. "Household Risk Management and Optimal Mortgage Choice." *Quarterly Journal of Economics* 118 (4): 1449–1494.

Carroll, Christopher D, Misuzu Otsuka, and Jiri Slacalek. 2011. "How Large Is the Housing Wealth Effect? A New Approach." *Journal of Money, Credit and Banking* 43 (1): 55–79.

Case, Karl E, John M Quigley, and Robert J Shiller. 2005. "Comparing Wealth Effects: The Stock Market Versus the Housing Market." *Advances in Macroeconomics* 5, no. 1.

Chen, Hui, Michael Michaux, and Nikolai Roussanov. 2020. "Houses as ATMs: mortgage refinancing and macroeconomic uncertainty." *The Journal of Finance* 75 (1): 323–375.

Chetty, Raj, László Sándor, and Adam Szeidl. 2017. "The effect of housing on portfolio choice." *The Journal of Finance* 72 (3): 1171–1212.

Choi, Hyun-Soo, Harrison Hong, and Jose Scheinkman. 2014. "Speculating on home improvements." *Journal of Financial Economics* 111 (3): 609–624.

Cocco, Joao F. 2005. "Portfolio Choice in the Presence of Housing." *Review of Financial Studies* 18 (2): 535–567.

Corradin, Stefano, Jose L. Fillat, and Carles Vergara-Alert. 2014. "Optimal Portfolio Choice with Predictability in House Prices and Transaction Costs." *Review of Financial Studies* 27 (3): 823–880.

Dynan, Karen. 2012. "Is a Household Debt Overhang Holding Back Consumption?" *Brookings Papers on Economic Activity*, pp. 299–362.

Fischer, Marcel, and Michael Z. Stamos. 2013. "Optimal Life Cycle Portfolio Choice with Housing Market Cycles." *Review of Financial Studies* 26 (9): 2311–2352.

Flavin, Marjorie, and Takashi Yamashita. 2002. "Owner-Occupied Housing and the Composition of the Household Portfolio." *American Economic Review* 92 (1): 345–362.

Follain, James R., and Stephen Malpezzi. 1981. "Are Occupants Accurate Appraisers?" *Review of Public Data Use* 9 (1): 47–55.

Gargano, Antonio, Marco Giacoletti, and Elvis Jarnecic. 2023. "Local experiences, search, and spillovers in the housing market." *The Journal of Finance* 78 (2): 1015–1053.

Goodman Jr., John L., and John B. Ittner. 1992. "The Accuracy of Home Owners' Estimates of House Value." *Journal of Housing Economics* 2 (4): 339–357.

Guren, Adam M, Alisdair McKay, Emi Nakamura, and Jn Steinsson. 2020. "Housing Wealth Effects: The Long View." *The Review of Economic Studies* 88 (2): 669–707 (04).

Hall, Robert E., and Frederic S. Mishkin. 1982. "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households." *Econometrica* 50 (2): 461–481 (March).

Ihlanfeldt, Keith R., and Jorge Martínez-Vázquez. 1986. "Alternative Value Estimates of Owner-Occupied Housing: Evidence on Sample Selection Bias and Systematic Errors." *Journal of Urban Economics* 20 (3): 356–369.

Kain, John F., and John M. Quigley. 1972. "Note on Owner'S Estimate of Housing Value." *Journal of American Statistical Association* 67 (340): 803–806.

Kaplan, Greg, Kurt Mitman, and Giovanni L Violante. 2020. "The housing boom and bust: Model meets evidence." *Journal of Political Economy* 128 (9): 3285–3345.

Kiel, Katherine A., and Jeffrey E. Zabel. 1999. "The Accuracy of Owner-Provided House Values: The 1978–1991 American Housing Survey." *Real Estate Economics* 27 (2): 263–298.

Kish, Leslie, and John B. Lansing. 1954. "Response Errors in Estimating the Value of Homes." *Journal of American Statistical Association* 49 (267): 520–538.

Kuchler, Theresa, Monika Piazzesi, and Johannes Stroebel. 2023. "Housing market expectations." In *Handbook of economic expectations*, 163–191. Elsevier.

Landvoigt, Tim. 2017. "Housing Demand During the Boom: The Role of Expectations and Credit Constraints." *Review of Financial Studies* 30 (6): 1865–1902.

Li, Geng, Kerwin Kofi Charles, Sheldon Danziger, and Robert F. Schoeni. 2007, April. “Studying Consumption with the Panel Study of Income Dynamics: Comparisons with the Consumer Expenditure Survey and an Application to the Intergenerational Transmission of WellBeing.” Finance and economics discussion series 2007-16, Board of Governors of the Federal Reserve System. Earlier version circulated in 2006.

Mian, Atif, and Amir Sufi. 2011. “House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis.” *The American Economic Review* 101 (5): 2132–2156.

Poterba, James M. 1984. “Tax Subsidies to Owner-Occupied Housing: An Asset-Market Approach.” *The Quarterly Journal of Economics* 99 (4): 729–752.

Robins, Philip K., and Richard W. West. 1977. “Measurement Errors in the Estimation of Home Value.” *Journal of American Statistical Association* 72 (358): 290–294.

Saiz, Albert. 2010. “The Geographic Determinants of Housing Supply*.” *The Quarterly Journal of Economics* 125 (3): 1253–1296 (08).

Yao, Rui, and Harold H. Zhang. 2005. “Optimal Consumption and Portfolio Choices with Risky Housing and Borrowing Constraints.” *Review of Financial Studies* 18 (1): 197–239.

Appendix

A-I Further Analysis of Misestimation

Socioeconomic effects of misestimation were investigated through regression analyses. These models explored the relationships between misestimation and variables such as family income, family size, gender, education, marital status, employment, and tenure. Fixed effects at the 5-digit ZIP code level, year, and age were included, and clustered standard errors at the ZIP, year, and age levels were used for robust inference.

The determinants of changes in house value misestimation were assessed through six regression models. Model 1 examined the relationship between changes in misestimation, changes in house value, lagged misestimation, and other controls. Model 2 focused on undervaluers (cases where misestimation was less than zero), while Model 3 analyzed overvaluers (cases where misestimation was greater than or equal to zero). Model 4 investigated changes in market house values as a determinant of misestimation for overvaluers, and Model 5 applied the same approach to undervaluers. Model 6 expanded on the analysis of overvaluers with market value changes. Across all models, fixed effects at the ZIP, year, and age levels were applied, and standard errors were clustered at these levels for statistical robustness.

Additional analyses included Ordinary Least Squares (OLS) and Instrumental Variable (IV) models to examine key financial behaviors. These models explored relationships such as stock value over liquid wealth, food consumption as a fraction of liquid wealth, stock participation and value for stockholders, risk-free assets over total wealth, and home improvements over market house value for stockholders. Extreme values were controlled by winsorizing variables at the 5th and 95th percentiles within each state and trimming the data.

To enhance model precision, several controls were incorporated, including mortgage size, house value, and local socioeconomic factors such as family size, family income (log-transformed), gender, education, and marital status. Fixed effects for state, year, and age were included, and when Google Trends data was used as an instrumental variable, DMA codes were added as fixed effects. A range of instrumental variables was also employed: elasticity data from Saiz, gamma variables from Guren, transaction bins from CoreLogic, and decile bins from Google Trends. These elements provided a robust framework for understanding the nuanced interactions between house value misestimation, financial variables, and socioeconomic trends.

Table A-1 provides an overview of the socioeconomic characteristics of the households included in our analysis, such as family income, household size, age, gender, education, marital status, and employment status, which serve as control variables in our empirical analysis. The sample is restricted to households with liquid wealth exceeding $2,963.3, corresponding to the average monthly salary in the U.S. in 2000. The table reports the mean, standard deviation, 5th and 95th percentiles, and the total number of observations for each variable.

Table A-1: **Descriptive Statistics of Socioeconomic and Instrument Variables.** This table presents summary statistics for the primary variables in the analysis, including the sample mean, standard deviation, 5th and 95th percentiles, and the total number of observations. Housing market data are sourced from the Federal Housing Finance Agency (FHFA), while data on house value misestimation, household socioeconomic characteristics, and financial decisions are obtained from the Panel Study of Income Dynamics (PSID). The sample is restricted to households with liquid wealth exceeding \$2963.3 ($LW > 2963.3$), which corresponds to the average monthly salary in the U.S. in the year 2000.

	Mean	Std. Dev.	p5	p95	Obs.
Socioeconomic Characteristics:					
Family Income (log)	10.885	0.959	9.278	12.223	60,373
Family Size	3.043	1.419	1.000	5.000	60,901
Age	43.950	14.347	25.000	72.000	60,893
Gender (Male=1)	0.872	0.334	0.000	1.000	60,900
Education (High School or More = 1)	0.396	0.489	0.000	1.000	59,128
Married (Married=1)	0.792	0.406	0.000	1.000	60,897
Employed (Employed=1)	0.840	0.367	0.000	1.000	60,876
Tenure	5.706	6.303	1.000	19.000	60,901
Instrument Variables:					
Elasticity FHFA, ε_{HPI}	238.453	170.215	54.785	585.030	38,848
Elasticity FHFA year of purchase, $\varepsilon_{HPI} \times HPI_{t_0}$	203.844	153.071	37.453	510.682	38,848
γ_{HPI}	132.957	91.916	39.300	305.415	30,111
$\gamma_{HPI} \times HPI_{t_0}$	112.242	80.656	32.221	268.431	30,111
Transaction 10 Bins	6.267	2.584	2.000	10.000	46,827
Google Trends Bins	5.452	2.619	1.000	10.000	25,437

House value misestimation may simply vary because homeowners learn and adjust the subjective valuation of their house in the direction of its market value or just because market values move, with no changes in subjective valuations. The latter would be consistent with homeowners anchoring their house value at the purchase price, for example. Specifically, learning and anchoring could play a significant role in explaining the variation of misestimation within ZIP codes, across households. In Table A-2 we show the results of OLS panel regressions of changes in house value misestimation at the household level for different subsamples on these potential determinants of misestimation dynamics. We control for lagged levels of misestimation, the current subjective value of the house (in logs), and the socioeconomic characteristics described in the previous table.

We find that both changes in the subjective and market valuations of the house have a significant impact on misestimation: misestimation increases when the household increase the subjective value, and misestimation decreases when market values grow, all else equal. The coefficients for ΔHV^S and ΔHV^M are positive and negative, respectively, which confirms that the variation in misestimation is driven by variation in both subjective and market valuation growth rates. These results are not driven by the subsample of households that undervalue or by the subsample of

Table A-2: **The Determinants of Changes in House Value Misestimation.** This table shows analysis of the determinants of house value misestimation. The dependent variable for all specifications is the change in misestimation (in \$100,000), Δm_{it}. The independent variables are the change in the subjective house value, ΔHV^S, and the change in the market house value, ΔHV^M. We control for lagged misestimation, $m_{i,t-1}$ and the house value in terms of HV^S (log). Our set of controls include the logarithm of family income, education (high school or more=1), employment status of the head of the household (employed=1), and the number of family members. We also control for the age, gender (male=1), marital status (married=1), tenure, tenure squared, and the number of transactions in the household's ZIP code. All our estimations use year and ZIP code-level fixed effects. Robust t-statistics are shown in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level of confidence, respectively. Standard errors are clustered at the year and ZIP code level.

	(1) All Households	(2) Only Undervaluers	(3) Only Overvaluers	(4) All Households	(5) Only Undervaluers	(6) Only Overvaluers
ΔHV^S	0.574***	0.543***	0.584***			
	(0.0491)	(0.0345)	(0.0487)			
ΔHV^M				-0.357***	-0.502***	-0.219***
				(0.0621)	(0.0534)	(0.0484)
$m_{i,t-1}$	-0.126***	-0.102*	-0.208***	-0.273***	-0.170***	-0.423***
	(0.0271)	(0.0537)	(0.0315)	(0.0380)	(0.0331)	(0.0937)
HV^S(log)	-0.0608*	-0.0706*	-0.0341*	0.265***	0.284***	0.289***
	(0.0305)	(0.0371)	(0.0171)	(0.0356)	(0.0355)	(0.0438)
Observations	43,789	18,304	24,671	43,789	18,304	24,671
R-squared	0.493	0.417	0.613	0.234	0.364	0.273
ZIP FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes

households that overvalue, as shown in columns (2)-(3) and (5)-(6).

All our specifications in Table A-2 include lagged misestimation, $m_{i,t-1}$. The negative term indicates that a higher level of misestimation is related to smaller future increases of misestimation. Households tend to react, as a higher misestimation triggers a decline in future misestimation, everything else equal, suggesting that there is mean reversion in misestimation.

Overall, we find that both growth in the subjective house value (i.e., actively updating the subjective valuation) and growth in the market value (i.e., the subjective value being sticky) play a significant role in explaining the variation of misestimation within ZIP codes, across households.

A-II Correlation Between Housing and Stock Returns

In this appendix, we justify our assumption of zero correlation between housing and stock returns. Figure A-1 visually illustrates the absence of a relationship between U.S. housing returns and U.S. stock returns over the period analyzed in our empirical study.

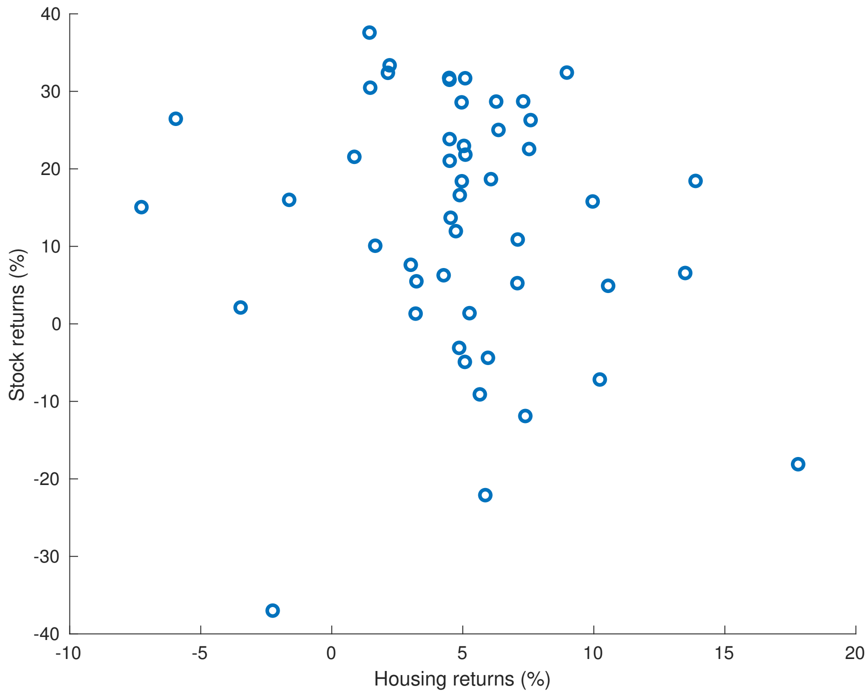

Figure A-1: **Historical US Stock Returns vs US Housing Returns.** This figure displays the scatter plot of US Stock Returns and US Housing Returns for the period 1984-2024.

Online Appendix

OA-I Data Description

The Panel Study of Income Dynamics (PSID) is a longitudinal household-level survey that began in 1968, tracking households and their offspring over time. With over five decades of data collection, the PSID has recorded insights from more than 84,000 individuals, offering an invaluable resource for understanding household dynamics and economic trends across generations. This study leverages the PSID to analyze the misestimation of household values, focusing specifically on periods when households move. At the time of relocation, the market value of the property is known, providing a reliable baseline for subsequent analyses of value misestimation. Using the Federal Housing Finance Agencys (FHFA) House Price Index (HPI) at the 5-digit ZIP code level, the dataset includes 20,769 observations of household moves and nearly 40,125 related observations of misestimation.

To create the database, several key steps were undertaken. First, individual-level data was processed by executing the IND2021ER .do file, which resulted in a dataset containing 84,121 records and 2,605 variables. This file was saved in .dta format for further analysis. Family-level data, spanning the years 1968 to 2021, was processed next. This dataset includes one record for every family interviewed since 1968, even if they participated in only a single wave of the survey. For each year, .do files were executed and the outputs saved as .dta files. Since all files were initially stored in a shared folder, scripts were adapted to facilitate their transfer and organization into a more manageable structure.

Geocode data covering census tracts, states, ZIP codes, Core-Based Statistical Areas (CBSAs), and Metropolitan Statistical Areas (MSAs) from 1968 to 2021 was also processed. This dataset comprises 306,189 observations, and scripts were modified to unify these files into a cohesive format suitable for longitudinal analysis. Additionally, wealth data collected during selected years1984, 1989, 1994, 1999, 2001, 2003, 2005, and 2007was incorporated. The wealth files were processed by modifying file paths in the original scripts, resulting in a consolidated dataset focused on wealth-related variables that are crucial for understanding economic trajectories.

Raw data from various external sources was integrated to enhance the dataset. This included FHFA HPI data at the 5-digit ZIP code level, annual house price indices adjusted to a 1990 benchmark, and Google Trends data, which was merged with DMA codes and ZIP codes. Once compiled, the raw data underwent a rigorous process of standardization, cleaning, and merging. Variable names and formats were standardized across all years, and identifiers were added where necessary. Observations with excluded valuessuch as house prices labeled 9999998 or 9999999 were removed to ensure data accuracy. The data was reshaped from a wide format, where variables for different years were stored across columns, to a long format, where each observation represents a unique household-year combination. This process resulted in a comprehensive dataset with 1,411,432 observations and 95 variables.

To create a unified dataset, heads of households were extracted from the individual-level data and saved in temporary files. ZIP code variables were integrated from the geocode files by matching FamilyID and year. These were merged with family and individual-level datasets, producing a consolidated dataset with 32,824 unique head-of-household observations across all years and 1,661 variables. The wealth datasets were similarly integrated. Variables from the selected years were cleaned, renamed, and merged with the family-level data in a one-to-one relationship, enabling robust longitudinal analyses that consolidate wealth variables with family identifiers.

The 05 Merge process involved multiple steps to integrate and enrich the dataset with key variables. First, the HPI ZIP Code Data from FHFA was prepared by creating a two-year growth variable for HPI growth, calculated using the formula:

$$\text{Two-Year HPI Growth (\%)} = \left[\left(1 + \frac{\text{Previous Year HPI}}{100}\right) \cdot \left(1 + \frac{\text{Current Year HPI}}{100}\right) - 1\right] \cdot 100$$

This same formula was applied to create a two-year growth variable for state-level HPI. Subsequently, ZIP code-level HPI data was merged in a many-to-one relationship, resulting in 1,411,432 observations and 90 variables, while state-level HPI data brought the dataset to 95 variables. Additional integrations included elasticity data (Saiz) by ZIP code (106 variables), FHFA index data by year (107 variables), and CBSA crosswalk and gamma data (Guren) by ZIP code and CBSA (111 variables). CoreLogic ZIP code liquidity data was merged by year and ZIP code, expanding the dataset to 112 variables, while Google Trends data added further detail, culminating in 116 variables. The process also computed elasticity using the state HPI index (Saiz) and created a variable for elasticity and year of purchase, assigning elasticity values to movers, thereby enriching the dataset for comprehensive longitudinal analysis.

The refine the dataset and prepare it for analysis. First, observations from the year 1968 were removed, dropping 32,824 entries, followed by the removal of records with missing ZIP codes, eliminating an additional 161,926 observations. Key control variables were then created or renamed for regression purposes, including Family Income (log-transformed), Gender (coded as 1 for male), Education (coded as 1 for college-educated individuals), Marital Status (coded as 1 for married individuals), and Employment (coded as 1 for individuals employed full- or part-time). Variables to track household movement and the year of purchase were introduced, along with a calculation for the gamma HPI (Guren) using the FHFA5 index, with adjustments made for movers and carried forward within families.

To estimate the market value of houses over time, we implemented a stepwise procedure to address missing values and account for annual appreciation rates. The variable HI, representing home improvements, was initialized to zero for observations with missing values:

$$\text{HI} = 0, \quad \text{if HI is missing.}$$

Next, we calculated the adjusted house value, HV_M_HI, which incorporates both the baseline house value (HV) and any home improvements (HI). For households that moved (move = 1), the value of HV_M_HI was set equal to HV:

$$\text{HV\_M\_HI} = \text{HV}, \quad \text{if move} = 1.$$

For households that did not move (move = 0), we estimated HV_M_HI iteratively using prior-year data. The value was updated by applying annual appreciation rates from the housing price index (HPI), specifically HPI_FHFA5 for years before 1999 and HPI_FHFA5_2years for years after 1997. The updates were computed as follows:

1. For years before 1999:

$$\text{HV\_M\_HI}_t = \text{HV\_M\_HI}_{t-1} \times \left(1 + \frac{\text{HPI\_FHFA5}}{100}\right) + \text{HI}, \quad \text{if HV\_M\_HI is missing.}$$

2. For years after 1997:

$$\text{HV\_M\_HI}_t = \text{HV\_M\_HI}_{t-1} \times \left(1 + \frac{\text{HPI\_FHFA5\_2years}}{100}\right) + \text{HI}, \quad \text{if HV\_M\_HI is missing.}$$

These steps ensure that the adjusted house value reflects both home improvements and cumulative appreciation, while accommodating differences in appreciation rates over time. The iterative approach accounts for the temporal dependency of house values, ensuring consistency in the panel dataset.

Financial variables were also meticulously prepared. Stock value was calculated as the sum of publicly traded stock holdings, mutual funds, or investment trusts. Checking and savings accounts combined balances across checking, savings, and money market accounts. Other assets encompassed life insurance cash values, valuable collections, or rights in trusts or estates. Cash assets included certificates of deposit, government bonds, and treasury bills, while debt was defined as the sum of all credit card, student loan, medical bill, legal bill, and family loan liabilities. Liquid wealth was defined as the total of stock value, checking and savings accounts, other assets, and cash, minus debt (excluding IRAs). Risk-free assets were defined as the sum of checking and savings accounts, other assets, and cash. Total wealth aggregated the values of various asset types, including farm or business ownership, real estate (other than the main home), vehicles, and private annuities, net of debt and inclusive of home equity.

Additional variables were created to measure misestimation as the difference between subjective house value and market house value (inclusive of home improvements). Consumption variables were calculated as the sum of expenditures on food at home, food delivery, and dining out. Indicators for stock participation (coded as 1 for stockholders) and stock value for stockholders were introduced. The dataset was further enriched with lags, changes, and logarithmic transformations of key variables such as misestimation, subjective house value, market house value, mortgage, and food consumption.

To prepare the Instrument Variables, we defined the panel structure by specifying the unique family identifier (`family_id`) and the temporal variable (`year`).

This step ensures that subsequent calculations respect the panel structure of the data, treating observations as part of a time series within families.

Next, we constructed an instrumental variable (IV) to address potential endogeneity in the model. The variable `IV` is derived from `trans_10_bins`, which represents decile bins for a specific variable, and the sign of the lagged misestimation variable (`misper_100k`). The calculation was conducted separately for years before and after 1998:

$$\mathrm{IV}_t = \begin{cases} \text{trans\_10\_bins} \times \text{sign}(l1.\text{misper\_100k}), & \text{if year} < 1998, \\ \text{trans\_10\_bins} \times \text{sign}(l2.\text{misper\_100k}), & \text{if year} \geq 1998. \end{cases} \tag{OA-1}$$

Additionally, we generated dummy variables representing each decile bin (`Bin10_1` through `Bin10_10`). These were interacted with the lagged sign of `misper_100k` to create expressions as follows:

$$\text{Bin10}_i = \begin{cases} \text{trans\_10\_bins}_i \times \text{sign}(l1.\text{misper\_100k}), & \text{if year} < 1998, \\ \text{trans\_10\_bins}_i \times \text{sign}(l2.\text{misper\_100k}), & \text{if year} \geq 1998, \end{cases} \tag{OA-2}$$

for $i = 1, \ldots, 10$.

To assess the relationship between observed trends and the misestimation variable, we incorporated Google Trends data. Using decile bins for Google Trends (`google_trends_bins`), we generated interaction terms between these bins and the second lag of `misper_100k`:

$$\text{google\_bins}_i = \text{google\_trends\_bins}_i \times \text{sign}(l2.\text{misper\_100k}), \tag{OA-3}$$

for $i = 1, \ldots, 10$.

These transformations facilitated the exploration of nuanced interactions between misestimations, decile bins, and external trends in a structured panel data framework.

The process refined the dataset by addressing missing or extreme values and developing models to explore the determinants and effects of house value misestimation. The initial step involved removing all observations with missing misestimation values, resulting in a dataset of 60,194 observations (4.4% of the original dataset). Additionally, observations with negative stock values (13 entries) were dropped. To address outliers, key variables such as misestimation, house value, and mortgage were winsorized at the 1st and 99th percentiles within each year. Descriptive statistics for these variables are summarized in the accompanying table.

OA-II Google Trends

Google Trends is a public-web that analyze the popularity of a query across regions and over time. It has a historical search data dating back to 2004. It provides a normalized index of search volume data so you can explore trends. The data is presented on a scale fo 0 to 100, representing the relative search interest of a given query compared to the highest point over the selected region and time frame. Google Trends allow us to compare multiple (up to five) queries. We used Google Trends to understand the consumer behavior over time and across regions.

We created a dictionary of possible queries that individual could have search in google. For this, we ask ChatGPT "What 25 words are the most searched in Google and other search engines when people is trying ot buy or sell a house?". The following list is the answer of ChatGPT gave us: "Real estate near me", "Real estate for sale", "New homes", "Real estate listings", "Apartments for rent", "Houses for rent", "Houses for sale near me", "Houses for sale", "Land for sale near me", "Land for sale", "For sale by owner", "Realtor near me", "Vacation rentals", "Condos for sale", "New construction homes near me", "Selling a house", "Cost of selling a house", "How to sell my house", "Taxes on selling a house", "Capital gains on selling a house", "How much does it cost to sell a house", "Realtor", "Real estate agent", "Real estate agent near me", and "Top real estate agents". Additional, to complete the dictionary, we asked:"Top 25 search in google when a person is trying to sell or buy a house in the USA since 2004" The result was two list, one for buying and another for selling, we added to the next queries to the dictionary: "Homes for sale", "Mortgage calculator", "Home buying tips", "First-time homebuyer programs", "Best neighborhoods to buy a house", "Home affordability calculator", "Mortgage rates", "Home inspection checklist", "Home buying process", "Down payment assistance programs", "Homebuyer grants", "Closing costs for buyers", "Types of mortgages", "Home appraisal process", "Home warranty", "Property taxes by ZIP code", "Home insurance quotes", "Buying a house with bad credit", "Buying a foreclosure", "Home inspection tips", "Homebuyer seminars", "Buying a house checklist", "Real estate market trends", "Home value estimator", "Home staging tips", "Selling a house by owner (FSBO)", "Home selling process", "Best time to sell a house", "Selling a house with a real estate agent", "Home selling tips", "Pricing my house to sell", "Closing costs for sellers", "Disclosure requirements when selling a house", and "Selling a house with tenants". Moreover, we added "Zillow" because of the significance over time. We ended up with a dictionary of 54 different queries.

As Google Trends only allows you to compare 5 queries. We selected "Houses for sale" search as a reference query. We split the dictionary in 14 lists of 4 queries per list.

For each list, we add up the reference query. We computed the values of the group in the United States from 2005 to 2021, but obly even years. We calculated the mean of each year for each query because the data obtained is in a monthly basis. Then we calculated the interest by region per year per query with a resolution of Designated Market Area (DMA), there are 210 DMAs in the USA. The interest per region result was normalized by scaling it based on yearly values in the yearly database. We added each list to a new data-frame multiplying it by a factor with the reference query. We replace infinity and NaN values with zeros to prevent errors during further analysis. We finished with 1890 observations.

We cleaned the database by removing all columns that all values are zero. Leaving us with 30 queries, this told us that from the dictionary proposed 24 searchers where not very popular against the other-ones. Then calculated the mean per row, meaning that we calculated the mean per year and region.

Finally, to create the instrument variable. We groped the data by year and calculated the quantiles (10) for the Mean column within each year. We assigned each row to a bin based on its Mean value relative to the quantiles.

From the remaining queries Having the value of each DMA region per year, we group it by year and then divided it in deciles and assigned a number of each decile per year. We added a bins column to the DataFrame, indicating the bin assignment for each year and region.