The Mortgage-Cash Premium Puzzle *

Michael Reher† and Rossen Valkanov‡

January 2022

Abstract

We document that mortgaged homebuyers pay an 11% premium relative to all-cash buyers in residential real estate transactions. This premium far exceeds the 3% premium implied by a realistically calibrated model of rational home sellers with transaction frictions. We obtain similar results from various estimators (e.g., repeat-sales, instrumental-variable, matching, semi-structural), novel data on non-accepted offers, and an experimental survey of U.S. homeowners. Experimental evidence suggests that pessimistic priors and uncertainty driven by inexperienced sellers best explain the puzzle of 8% (11% − 3%). Our findings matter economically, as all-cash purchases account for one-third of all U.S. home purchases over 1980-2017.

Keywords: House Prices, Cash Buyers, Asset Pricing Puzzles, Affordability
JEL Classification: G31, R30, G12, G21, G41

*We thank On Amir, Carlos Avenancio-Leon, Paul Chiou, Andra Ghent, Sonia Gilbukh, Uri Gneezy, Lu Han, Jun Liu, Gregor Matvos, Peter Maxted, Will Mullins, Chris Parsons, Alberto Plazzi, Alessandro Previtero, Alberto Rossi, Yuval Rottenstreich, Gregor Schubert, Rick Townsend, Neng Wang and seminar participants at the SFS Cavalcade, EFA Annual Meeting, AREUEA, PBFEAM, NFA Annual Meeting, FMA Annual Meeting, UC San Diego, Santa Clara University, University of Wisconsin, and the European Winter Finance Summit for comments. Ziyi Chen provided excellent research assistance. We are grateful to the following data vendors for providing their data under academic terms: Zillow, Recursion Co, and the California Association of Realtors. We thank Redfin for the opportunity to analyze their data, and we especially thank Taylor Marr who facilitated the process. The views expressed in this paper are those of the authors and do not reflect the position of Zillow, Recursion, the California Association of Realtors, or Redfin. This study is exempt from Institutional Review Board requirements under category 45 CFR 46.104(d)(2).

†University of California San Diego, Rady School of Management. Email: mreher@ucsd.edu
‡University of California San Diego, Rady School of Management. Email: rvalkanov@ucsd.edu

1 Introduction

Consider a home seller with offers from two competing buyers: one is mortgage-financed, and the other is all-cash. Define the mortgage-cash premium as the expected difference in log prices between the two offers. In the absence of frictions, the mortgage-cash premium ought to be zero (Modigliani and Miller 1958). More realistically, the premium ought to be positive to compensate sellers for frictions in the mortgage origination process, namely risk of transaction failure and a longer time to close. We find that the mortgage-cash premium averages 11% over the past 40 years. In dollar terms, a seller would be indifferent between a \$500,000 all-cash offer, the average in our sample, and a mortgaged offer that is \$55,000 larger. The magnitude of this premium far exceeds reasonable compensation for transaction frictions implied by traditional economic logic. In policy terms, U.S. taxpayers subsidize \$8 trillion of mortgages to promote homeownership (Federal Reserve 2019); reducing the mortgage-cash premium would enable a smaller subsidy to accomplish the same goal.

To make things concrete, consider the following example. A risk-neutral home seller decides whether to accept a riskless, cash-financed purchase offer versus a mortgage-financed offer that fails with 6% probability (NAR 2020). If the transaction fails, the seller relists the home in one month after reducing the price by 10%, equal to the average price cut during the 2008 financial crisis (Trulia 2009). Supposing the seller is indifferent and has a discount rate of zero over this short horizon, the mortgage-financed offer should require a 0.6% price premium, which obtains from solving

$$0 = \underbrace{94\%}_{\text{Success Rate}} \times \textit{Premium} - \underbrace{6\%}_{\text{Failure Rate}} \times \underbrace{10\%}_{\text{Cost of Failure}} . \tag{1}$$

Instead, we estimate a premium of 11%, which is robust to a variety of datasets and identification strategies. Even a more realistic model with risk aversion, time discounting, and additional frictions still only implies a premium of 3%. The remaining 8% constitutes a puzzle from the standpoint of traditional models. Turning to models in which nontraditional channels amplify these frictions, we find that heterogeneous beliefs and uncertainty aversion best explain the large mortgage-cash premium.

In more detail, we begin by using a variety of datasets to document that cash-financed purchases account for one-third of all U.S. home purchases over the past 40 years. This fact is surprising given that the literature tends to focus on mortgage-financed purchases. We then document that the average cash-financed purchase is associated with a 30% lower price relative to the average mortgage-financed purchase. At a descriptive level, these facts suggest that sellers prefer cash financing and, accordingly, require mortgage-financed buyers to pay a premium. But, of course, we cannot reach this conclusion on the basis of summary statistics alone. For example, cheap properties may attract all-cash buyers, which is the reverse of the causal chain we seek to estimate.

We devote the majority of the paper to estimating the premium that mortgaged buyers must pay to purchase the same home relative to all-cash buyers: the mortgage-cash premium. This exercise is challenging because we do not observe a property's counterfactual sales price under the alternative method of financing. We take a transparent approach to this challenge, using a large dataset on U.S. home purchases to estimate a combined repeat-sales and hedonic pricing equation. This approach effectively compares the same property sold at different times under different methods of financing. Its validity depends on the ability of the controls and fixed effects to absorb enough variation such that any two purchases are as-good-as-equal, up to the method of financing. Our data's breadth allows us to include such a large set of controls and fixed effects that they collectively explain over 90% of the variation in sales prices. Consequently, there is little scope for bias based on unobserved features of the purchase.

We estimate a mortgage-cash premium of 11.7% using this repeat-sales-hedonic estimator. We obtain similar estimates when: excluding homes that are subsequently flipped; excluding transactions involving an institution, a non-U.S. party, or an underwater seller; or when restricting the sample to newly built homes. This last filter implies that the estimated premium does not confound a premium for adverse selection, since, empirically, almost all new construction is of high-quality housing (e.g., Rosenthal 2014). Consistent with the result's external validity, we estimate a similar premium of 12.2% based on a purchase-level dataset from an entirely different provider (CoreLogic) than our baseline provider (Zillow).

Next, we take a serious and systematic approach to the question of selection bias. Such bias can take four forms: internal invalidity due to correlation between the method of financing and price-relevant characteristics of the buyer, the seller, or the property's condition; or external invalidity. Based on five different datasets and ten different estimators designed to assess the scope for selection bias, we consistently estimate a mortgage-cash premium between 8.6% and 16.9%, straddling our baseline estimate of around 11%.

We address bias from buyer characteristics by using novel data on non-accepted, off-equilibrium offers to construct a counterfactual. Intuitively, this approach calculates the difference-in-difference between winning versus losing offers between mortgaged versus cash-financed purchases. Thus, we can separate the mortgage-cash premium from the propensity of mortgaged buyers to make overpriced offers because of, say, informational disadvantages or lack of negotiation skill. We estimate an 8.6% premium using this offer-level approach, which, given its minimal identification assumptions, provides a credible lower bound.

To address bias from seller characteristics, we return to our baseline dataset and control for a battery of variables that govern the seller's joint motivation to list at a low price and to prioritize all-cash offers (e.g., debt overhang, moving propensity) and the buyer's ability to identify such sellers (e.g., institutional status, financing in other transactions). Then, we identify the mortgage-cash premium using an instrumental variable: the share of other homes that the seller has sold to all-cash buyers. This instrument reflects the persistent component of the seller's cash preference, and so it avoids bias from temporary urgency that may jointly affect the seller's list price and her preferred method of financing. We estimate a 13.9% premium using this instrument. Lastly, we use listing-level data from a major state realtor association to control for the seller's list price, which gives a 14.3% premium.

The exercises designed to address bias from buyer and seller characteristics simultaneously address many issues related to time-varying property characteristics, such as the propensity of all-cash buyers to purchase properties in poor condition. We further investigate the scope for such bias through the Bajari et al. (2012) semi-structural estimator, which reduces bias by introducing a Markov structure for property condition, and through propensity score matching (e.g., Abadie and Imbens 2006), which restricts the comparison to highly similar transactions within the same zip code and year. These exercises give premiums of 14.9% and 16.9%, respectively.

Lastly, we assess whether the sampling restrictions required for internal validity jeopardize the results' external validity. We estimate a premium of 16.9% when including properties that do not experience a repeat sale, consistent with the property fixed effects reducing bias in our baseline sample. Then, we estimate a premium of 10.3% when weighting our baseline transactions by their inverse probability of appearing in the baseline sample, according to observed characteristics (e.g., Solon, Haider and Wooldridge 2015). Along with the previous exercises, these findings support the internal and external validity of an 11% premium.

Recalling that mortgaged transactions expose the seller to additional frictions, we find that the estimated mortgage-cash premium co-moves with the degree of friction in both the time-series and the cross-section. In the time series, the estimated premium peaks over 2005-2010, during which the probability of mortgage transaction failure and the price cut associated with failure also peak. In the cross-section, we estimate a larger premium for buyers with greater risk of obtaining financing, measured by the mortgage application denial rate for either the surrounding market. Likewise, we estimate a larger premium for sellers with limited ability to undergo a long and risky transaction, such as those with large mortgage debt overhang or who purchase another home concurrently.

The relevant question, however, is whether the degree of transaction friction necessitates a mortgage-cash premium of the magnitude that we estimate. We pursue this question of magnitude by calibrating a framework of rational home sellers with traditional ingredients and realistic frictions: risk aversion, time discounting, lengthy mortgage closing periods, mortgage-specific closing costs, home search costs, and the seller's debt overhang and down payment on her next home. A reasonable parameterization based on long-run data implies a calibrated mortgage-cash premium of no more than 3%, compared to our estimated premium of 11%. In fact, even our most conservative estimate of around 8% requires a parameterization more extreme than that associated with the 2008 financial crisis.

Given the stability of our empirical results across the exhaustive set of exercises described above, we ask whether our theoretical results under-predict the mortgage-cash premium because they rely only on highly traditional channels. We consider two sets of nontraditional channels: beliefs about mortgaged transactions that differ from long-run data; and distortions of these beliefs due to nonstandard preferences.

We assess these nontraditional channels by posing our motivating thought experiment to a survey of 2,000 geographically representative U.S. homeowners, administered in two waves. The survey reproduces our main finding: the average homeowner requires a 10.4% premium to accept a mortgaged offer over a competing all-cash offer. Importantly, this experimental premium is similar across the survey's first (10.2%) and second waves (10.5%), strongly supporting its validity and the validity of our baseline results. Following standard practice, we infer the experimental premium through a multiple price list presented in dollars and percent.

Consistent with a role for belief heterogeneity, the average respondent's prior probability of mortgage transaction failure equals 13.3%, compared to the contemporaneous failure rate of 7% based on market data. After accounting for such pessimistic priors, the gap between the average respondent's experimental and theoretical premium falls from 8.9 pps to 7.2 pps, or 1.7 pps (19%).

Next, we evaluate whether nonstandard preferences distort these beliefs by calculating the probability distortions implied by uncertainty aversion (e.g., Hansen and Sargent 2011), disappointment aversion (e.g., Gul 1991), and loss aversion (e.g., Kőszegi and Rabin 2006), as well as by probability weighting (e.g., Tversky and Kahneman 1992). Each distortion has an associated behavioral parameter with a surrounding literature dedicated to quantifying that parameter. Adhering to each literature as closely as we can, we find that uncertainty aversion best explains the experimental mortgage-cash premium. Geometrically, uncertainty aversion fits the data well because it matches the curvature in the relationship between the beliefs implied by a respondent's premium and her prior beliefs. Intuitively, this curvature reflects how aversion to uncertainty leads sellers to optimize according to a similar, worst-case probability of failure. By contrast, disappointment and loss aversion can only fit the data under unrealistic parameterizations, and probability weighting can only fit sellers with extremely optimistic beliefs.

Two pieces of non-structural evidence also support the importance of uncertainty. First, survey respondents who are told the distribution of mortgage transaction outcomes require a 1.4 pps lower premium than those who hold similar beliefs but face uncertainty. Second, in both our experimental and observational data, we estimate that the mortgage-cash premium falls by 0.7 pps for each time the seller has sold a home. Since sale experience plausibly reduces uncertainty, this finding further supports uncertainty aversion as the most promising probability distortion.

From a policy perspective, our results suggest that a seemingly modest easing of transaction frictions can have outsized effects on reducing the premium paid by mortgaged buyers. This conclusion implies that an easing of such frictions may be a more cost-effective route to promoting homeownership than subsidizing mortgages for first-time homebuyers.

We conclude the introduction by situating our contribution within the literature. Section 2 describes our data. Section 3 presents motivating facts. Section 4 estimates the mortgage-cash premium. Section 5 assesses selection bias. Section 6 calibrates the premium using a traditional framework. Section 7 discusses experimental evidence and nontraditional channels. Section 8 concludes.

Related Literature

First, we show how capital structure affects valuation in a fresh setting, the residential real estate market. Viewing the mortgage-cash premium as analogous to a "credit spread", we parallel a literature on how traditional models struggle to explain large credit spreads in bond markets.[1] Our conclusion differs, however, in that amplification from covariance between the probability and costs of transaction failure (i.e., "default") cannot explain the mortgage-cash premium. Instead, the most compelling explanation combines heterogeneous beliefs (e.g., Savage 1954) and belief distortions, the latter of which contributes to an analogous puzzle in the health insurance market (e.g., Barseghyan, Prince and Teitelbaum 2011; Einav et al. 2012; Barseghyan et al. 2013). In particular, the evidence supports uncertainty

[1] See, for example, Huang and Huang (2012), Chen (2010), Chen, Collin-Dufresne and Goldstein (2009), or Almeida and Philippon (2007).

aversion as the most relevant belief distortion.[2] This conclusion resembles Caskey (2008), who shows how combining uncertainty aversion with heterogeneous beliefs contributes to the equity premium in the stock market.

Second, by rigorously quantifying home sellers' preference for cash-financed purchases, we provide an alternative perspective on how credit markets affect house prices. This channel operates through frictions in the microstructure of real estate purchases and so complements channels that operate through the overall demand for housing, on which there is a large literature summarized by Glaeser and Sinai (2013) and Piazzesi and Schneider (2016). We find, however, that these frictions are "overpriced", thus offering a novel example of mispricing in real estate.[3] By quantifying the importance of such frictions, we also support a literature on the behavior of real estate agents who, in principle, help mitigate them.[4] Lastly, our particular focus on sellers' behavior after listing their home complements the Guren (2018) model of how sellers determine prices before listing.

Third, we contribute to an emerging literature on the role of all-cash buyers in real estate.[5] In this vein, we relate most closely to contemporaneous and independent work by Han and Hong (2020), Seo, Holmes and Lee (2021), and Buchak et al. (2020), who, respectively, document price discounts for all-cash buyers in Los Angeles, in Tallahassee, and for a specific type of all-cash buyer, iBuyers. We differ from these papers by estimating this price discount over a broader sample of all-cash buyers, but our main contribution lies in evaluating whether traditional channels can explain the discount's magnitude. After accounting for differences in sample, our estimated mortgage-cash premium agrees with the estimates in these papers.

[2] We model uncertainty aversion following the literature on robust decision-making (e.g., Cagetti et al. 2002; Anderson, Hansen and Sargent 2003; Maenhout 2004; Hansen and Sargent 2011; Barnett, Buchak and Yannelis 2021), which, as Hansen et al. (2002) show, can be isomorphic to another common treatment of uncertainty aversion proposed by Gilboa and Schmeidler (1989).

[3] This example complements others such as over-optimism about price growth (e.g., Kaplan, Mitman and Violante 2020; Foote, Loewenstein and Willen 2020; Glaeser and Nathanson 2017; Chinco and Mayer 2016; Cheng, Raina and Xiong 2014; Shiller 2014).

[4] See Barwick and Pathak (2015), Han and Hong (2011), Hsieh and Moretti (2003), Levitt and Syverson (2008), or Gilbukh and Goldsmith-Pinkham (2019) for examples.

[5] Early work by Asabere, Huffman and Mehdian (1992) and Lusht and Hansz (1994) find premiums of 13% and 16%, respectively, based on small samples of 300 purchases in two townships in Pennsylvania.

2 Data

Our analysis relies on five observational datasets and one experimental dataset, the last of which we describe in Section 7. Table 1 summarizes key variables from the two most important observational datasets, which we now describe. Appendix A has complete details.

The first and primary dataset is Zillow's Assessor and Real Estate Database (ZTRAX). The ZTRAX dataset contains information about home purchase transactions over 1980-2017. Zillow collects the data from a combination of public records and proprietary sources. To avoid misleading comparisons that could bias the estimates, we impose a variety of important filters to exclude properties in foreclosure, intrafamily transfers, purchases with a very high or low sales price or leverage ratio, and various other extreme cases, leading to a filtered ZTRAX universe of 11,367,195 transactions. For computational convenience, we draw a 25% random sample of this universe and collapse the data into an unbalanced panel across properties i and months t. The resulting dataset spans 80% of U.S. counties on a population-weighted basis (2,254,389 transactions). We prioritize internal validity and perform our baseline analysis on the subsample of properties for which we can include a property fixed effect, which spans 35% of counties (426,256 transactions). This restriction leads to conservative estimates, as we verify in Section 5.4.

The most important variables in the ZTRAX dataset are the sales price and loan-to-value (LTV) ratio associated with the purchase. We also observe the name of the seller and buyer involved in the transaction, which we use to perform several of the robustness exercises in Section 5 and to study heterogeneity in Section 6.4.3. Lastly, we observe various hedonic characteristics of the property, which, along with geographic and property fixed effects, help us achieve an R-squared above 90% in our baseline regression.

The second dataset contains information on both accepted and non-accepted purchase offers made through a large U.S. online real estate brokerage, Redfin. We use this offer-level dataset to assess the internal validity of the results obtained from the ZTRAX dataset in Section 5.1. The raw data are reported by real estate agents affiliated with Redfin and date back as far as 2013. Until recently, Redfin's real estate agent program covered only a few markets, and so we begin our analysis in January 2020 when the program crossed the threshold of covering 50% of counties.

The Redfin dataset includes information on the offer's price and method of financing, the geographic location of the property, the date on which the offer was made, the number of competing offers, and other variables described in Appendix A. Out of concern for client privacy, we cannot directly view the microdata and instead analyze the Redfin dataset by submitting a program with our desired calculations. We view this constraint as an advantage, as it limits the scope for data mining.

Lastly, we rely on four additional datasets to complete the robustness exercises in Section 5 and to calibrate the framework in Section 6. These include: a transaction-level dataset from CoreLogic that is analogous to ZTRAX; a survey of real estate agents conducted by the National Association of Realtors (NAR) over 2015-2021; mortgage application data from the Home Mortgage Disclosure Act (HMDA); and a survey of home sellers in California in 2019 conducted by the California Association of Realtors (CAR).

3 Motivating Facts

We motivate our empirical analysis with two facts about cash-financed home purchases. First, Figure 1a shows how such purchases account for 35% of all home purchases over 1980-2017, based on the ZTRAX dataset. Explicitly, we plot the distribution of loan-to-value (LTV) ratios, which features well-known bunching around various positive regulatory thresholds (e.g., Greenwald 2018). We focus on the less well-documented bunching that occurs at zero, corresponding to cash-financed purchases.[6]

Second, Figure 1b shows how the average cash-financed purchase over 1980-2017 has a lower real sales price relative to the average mortgage-financed purchase. Beginning with the leftmost column, the average difference in real sales price is 35 log points. Moving right, we

[6]We check that the ZTRAX dataset does not inflate the share of cash-financed purchases by underrepresenting mortgages originated by nonbank lenders. Explicitly, the share of mortgages originated by nonbanks in our baseline sample equals 40%, compared to 41% based on data from the Home Mortgage Disclosure Act (HMDA) and the definition of nonbanks in Gete and Reher (2021). The share of mortgage-financed purchases based on the CoreLogic dataset described in Appendix A equals 65%, consistent with ZTRAX.

restrict the sample to purchases in zip codes with bottom-quartile real income, top-quartile real income, and purchases in time periods without extreme house price fluctuations. Each partition gives a real price difference between 33 and 46 log points.

At a highly suggestive level, these two facts indicate that sellers prefer cash-financed buyers, per Figure 1a, and so they require mortgage-financed buyers to pay a "premium", per Figure 1b. In the next section, we rigorously estimate this mortgage-cash premium, focusing on its average value in the U.S. over the medium-to-long run.

4 Mortgage-Cash Premium

Our goal is to estimate the premium that mortgaged buyers must pay to purchase the same home relative to all-cash buyers: the "mortgage-cash premium". We formalize the econometric problem in Section 4.1 and present our baseline results in Section 4.2.

4.1 Identification

Consider an experiment in which we change the method of financing for a home purchase from mortgaged financing (i.e., debt-and-equity) to all-cash financing (i.e., only equity), holding all other aspects of the transaction fixed. We would like to compare the "shadow price" associated with each method of financing. In the notation of the potential outcomes literature (e.g., Rubin 1974), let $P_{i,t}^F$ denote the counterfactual sale price of property i in month t under method of financing $F \in \{M, C\}$, where M and C denote mortgaged and all-cash financing, respectively.

In equilibrium, we only observe the counterfactual price associated with the equilibrium method of financing. Explicitly,

$$\log\left(P_{i,t}\right) = \mathcal{M}_{i,t} \log\left(P_{i,t}^M\right) + \left(1 - \mathcal{M}_{i,t}\right) \log\left(P_{i,t}^C\right), \tag{2}$$

where $P_{i,t}$ is the observed sales price; and $\mathcal{M}_{i,t}$ indicates if the purchase is financed by a mortgage (i.e., $F = M$). Define the price premium paid by mortgaged buyers as

$$\mu \equiv \mathbb{E}\left[\log\left(P^{M}_{i,t}\right) - \log\left(P^{C}_{i,t}\right)\middle|\, \theta_{i,t}, \mathcal{M}_{i,t}\right] \tag{3}$$

where $\theta_{i,t}$ contains observed, price-relevant information. Identifying μ is challenging because a purchase can only have one method of financing in equilibrium, $\mathcal{M}_{i,t}$. Therefore, we do not observe $P^{M}_{i,t}$ for cash-financed purchases (i.e., $\mathcal{M}_{i,t} = 0$), and, similarly, we do not observe $P^{C}_{i,t}$ for mortgage-financed purchases (i.e., $\mathcal{M}_{i,t} = 1$).

We estimate equation (3) using a repeat-sales-hedonic approach in which the method of financing is treated as a time-varying "hedonic characteristic". This approach has the advantage of transparency and a long tradition in the literature (e.g., Giglio, Maggiori and Stroebel 2015). As its name implies, a repeat-sales-hedonic approach requires controlling for an exhaustive set of property fixed effects (i.e., "repeat sales"), observable property characteristics (i.e., "hedonic"), and other fixed effects. These ancillary parameters serve to absorb as much variation as possible, such that purchases only effectively differ in their price and their method of financing. The regression equation is

$$\log\left(Price_{i,t}\right) = \mu Mortgaged_{i,t} + \psi X_{i,t} + \zeta_{z(i),t} + \alpha_i + \epsilon_{i,t}, \tag{4}$$

where $Mortgaged_{i,t}$ indicates if the loan amount is positive (i.e., $\mathcal{M}_i$); $Price_{i,t}$ is the sales price (i.e., $P_{i,t}$); α_i is a property fixed effect; $\zeta_{z(i),t}$ is a zip code-by-month fixed effect; and $X_{i,t}$ is a vector of indicators for whether i belongs to bins defined by month and various hedonic characteristics.[7]

Equation (4) identifies the mortgage-cash premium, μ, under the following assumption,

$$\mathbb{E}[Mortgaged_{i,t} \times \epsilon_{i,t} |\, \underbrace{\alpha_i, \zeta_{z(i),t}, X_{i,t}}_{\theta_{i,t}}] = 0. \tag{5}$$

In words, assumption (5) states that, conditional on the various fixed effects and observed

[7]The hedonic characteristics are: the number of years from when the property was built; the number of overall rooms, bathrooms, and stories; and indicators for whether the property has air conditioning and is a detached single-family home. Standard errors are clustered by property.

characteristics (i.e., $\theta_{i,t}$), the method of financing (i.e., $Mortgaged_{i,t}$) does not correlate with unobserved, price-relevant characteristics of the transaction (i.e., $\epsilon_{i,t}$). To reiterate, the conditioning arguments in assumption (5) mean that equation (4) still identifies the mortgage-cash premium even if the method of financing correlates with time-invariant features through α_i (e.g., view); time-varying features of the local market through $\zeta_{z(i),t}$ (e.g., gentrification); and the potentially time-varying price of hedonic characteristics through $\psi X_{i,t}$ (e.g., value of space). We rigorously assess the validity of assumption (5) in Section 5.

4.2 Results

We estimate equation (4) using the ZTRAX dataset and report the results in Table 2. The estimated mortgage-cash premium over 1980-2017 equals 11.7%, as shown in column (1). The regression features an R-squared of 91%, largely due to the 287,000 property fixed effects. This high R-squared means that there is little remaining variation in unobserved characteristics to constitute a violation of our identification assumption (5). Indeed, consistent with the fixed effects reducing bias, we estimate a premium of 19.1% when only including zip code-by-month fixed effects in column (0). Section 5.4 verifies that the accompanying sample restriction does not jeopardize external validity.

The remaining columns of Table 2 report the estimated premium within various subsamples. In columns (2) through (5), we obtain a consistent premium after partitioning the sample by time period. The higher premium over 2005-2010 corresponds to elevated transaction frictions during that period, as we later discuss in Section 6.4.2.

Columns (6) and (7) restrict the sample to transactions with less asymmetric information about the property's condition. In column (6), we estimate a premium of 9.0% on properties built within the previous three years, the condition of which is likely quite good. Such properties are also less likely to possess antique features that appeal to a particular clientele, and so the estimate does not reflect a discount for the illiquidity of the niche property market. In column (7), we exclude transactions in which the buyer subsequently re-sells the property within 12 months. The remaining buyers in the sample are less likely to have uncovered a "lemon", and so the estimated premium of 8.8% does not suffer upward bias from an ex-ante informational discount required by these buyers.

In columns (8) and (9), we consider transactions between non-institutional and non-foreign parties, respectively. The respective premiums equal 11.2% and 10.8%. These findings imply that we do not confound how institutions transact in illiquid submarkets (e.g., Mills, Molloy and Zarutskie 2019) or how out-of-town buyers have informational disadvantages (e.g., Chinco and Mayer 2016).

Lastly, we estimate a premium of 9.9% after restricting the sample to transactions in which the seller likely has positive equity, as shown in column (10). Like in columns (6)-(7), such transactions have less asymmetric information about the property's condition.

Collectively, the stable estimates of between 8% and 12% in columns (6)-(10) support the internal validity of an 11% premium. These columns restrict the sample to relatively transparent purchases, and so, even if such purchases are less-likely to be financed with cash, such selection does not appear to bias the estimate.

5 Selection Bias

We take the question of selection bias very seriously, and so we perform over ten exercises using five different datasets to assess both internal and external validity. Table 3 concisely summarizes the resulting estimates of the mortgage-cash premium. The estimates range from 8.6% to 16.9%, straddling our baseline estimate (11.7%) and so supporting its validity.

Without loss of generality, we organize this section by expressing the pricing error $\epsilon_{i,t}$ as a weighted average of three innovations: variation in the property's condition that market participants observe but econometricians do not; the buyer's offer price relative to this condition; and, similarly, the seller's reservation value. Violations of internal validity occur when the average of these innovations covaries with the method of financing, contrary to assumption (5). Columns (6)-(10) of Table 2 already account for several specific violations of internal validity. In Sections 5.1, 5.2, and 5.3, we more systematically assess bias from buyer valuation, seller valuation, and property condition, respectively. In so doing, we impose sample restrictions that may affect external validity, which we assess in Section 5.4.

5.1 Buyer Offer Behavior

Mortgaged buyers may bid higher than all-cash buyers for three reasons. First, they may systematically hold higher private valuations of real estate because of, say, limited information or inflated expectations. Second, mortgaged buyers may hold similar private valuations as all-cash buyers, but they or their real estate agents lack the skill to successfully negotiate the price below this private valuation. These two channels violate assumption (5) because they do not stem from the method of financing. We seek to identify a third channel that indeed stems from the method of financing: mortgaged buyers offset financial frictions by offering a higher price (i.e., the mortgage-cash premium). We use an offer-level research design to separate this channel from the two confounding channels.

Using the notation from Section 4.1, let $P_{i,t}^{N}$ denote the price associated with a non-accepted offer to purchase property i in month t, and let $\pi_{i,t}^{M}$ and $\pi_{i,t}^{C}$ denote the log price premium paid by the winning offer under mortgaged and all-cash financing, respectively. That is, $\log\left(P_{i,t}^{M}\right) \equiv \pi_{i,t}^{M} + \log\left(P_{i,t}^{N}\right)$ and $\log\left(P_{i,t}^{C}\right) \equiv \pi_{i,t}^{C} + \log\left(P_{i,t}^{N}\right)$. We can then rewrite the mortgage-cash premium from equation (3) as

$$\mu \equiv \mathbb{E}[\underbrace{\left(\pi_{i,t}^{M} + \log\left(P_{i,t}^{N}\right)\right)}_{\log\left(P_{i,t}^{M}\right)} - \underbrace{\left(\pi_{i,t}^{C} + \log\left(P_{i,t}^{N}\right)\right)}_{\log\left(P_{i,t}^{C}\right)} |\theta_{i,t}, \mathcal{M}_{i,t}] = \mathbb{E}\left[\pi_{i,t}^{M} - \pi_{i,t}^{C} \middle| \theta_{i,t}, \mathcal{M}_{i,t}\right]. \qquad (6)$$

Equation (6) implies that the mortgage-cash premium equals the difference-in-difference between winning versus losing offers between mortgage versus cash-financed purchases. In particular, we difference out any confounding information encoded in non-accepted offers.

Accordingly, we estimate the following difference-in-difference regression equation,

$$\begin{aligned} \log\left(Price_{i,j,t}\right) = \mu \left(Mortgaged_{i,j,t} \times Winning_{i,j,t}\right) + ... \qquad (7) \\ ... + \psi_0 \, Winning_{i,j,t} + \psi_1 Mortgaged_{i,j,t} + \zeta_{z(i)} + \tau_t + \upsilon_{i,j,t}, \end{aligned}$$

where i, j, and t index property, offer, and month; $Mortgaged_{i,j,t}$ indicates if j is a mortgage-financed offer; $Winning_{i,j,t}$ indicates if j is the accepted offer; $Price_{i,j,t}$ is the price offered by

j; and $\zeta_{z(i)}$ and τ_t are zip code and month fixed effects.[8]

The identification assumption associated with equation (7) is

$$\mathbb{E}[Mortgaged_{i,j,t} \times Winning_{i,j,t} \times \upsilon_{i,j,t} | \zeta_{z(i)}, \tau_t, Mortgaged_{i,j,t}, Winning_{i,j,t}] = 0. \tag{8}$$

This assumption has two advantages over the baseline assumption (5). First, conditioning on $Mortgaged_{i,j,t}$ allows us to identify the mortgage-cash premium even if mortgaged buyers systematically hold higher private valuations and, thus, make higher offer prices. Second, conditioning on $Winning_{i,j,t}$ allows us to identify the premium even if all-cash buyers possess superior negotiation skill that enables them to win more often and at a lower price.

Table 4 reports the results. We estimate a premium of 8.1% in column (1). In column (2), we estimate a premium of 8.6% after controlling nonparametrically for competitiveness, as measured by the number of competing offers. These estimates lie close to those obtained on the subsample of non-flipped properties in the baseline Table 2, supporting the latter's internal validity.

It is important to stress how little we have assumed in this offer-level research design. We can, for example, identify the mortgage-cash premium even if all-cash buyers select different types of properties, hold different beliefs, or differentially win deals for non-financial reasons. Thus, the smallest estimate of around 8% from Table 4 provides a credible lower bound on the mortgage-cash premium.

5.2 Seller Reservation Value

Sellers with a low listing price, or, more generally, a low reservation value, may disproportionately sell to all-cash buyers for two reasons. First, all-cash buyers may have the skill required to identify such "motivated sellers". Second, motivated sellers may both list at a low price to attract more offers and also prioritize all-cash offers because they come with fewer frictions. This second case violates assumption (5) because sellers who strongly prefer

[8] We weight observations by the total number of offers on the listing to address the fact that we do not observe all offers. The unweighted estimates lie within a 1 pps bandwidth of the weighted estimates.

all-cash offers would nevertheless sell their home at a relatively low price even if an all-cash offer never arrived.

We address the first case by controlling for a battery of observed seller and buyer characteristics that proxy for the seller's motivation and the buyer's ability to infer such motivation. We address the second case by instrumenting for the method of financing using the seller's rate of acceptance of all-cash offers in other transactions. Thus, we identify the mortgage-cash premium through the persistent component of the seller's all-cash preference, rather than through transitory shocks that lead the seller to both list low and to accept all-cash offers (e.g., a health emergency).

5.2.1 Buyer and Seller Characteristics

We first reestimate equation (4) after controlling for observed characteristics of the seller and buyer.[9] The seller characteristics include indicators for whether the seller: has a loan-to-value ratio that exceeds 50%; purchases another home in the same month; and has a non-U.S. address. The buyer characteristics include indicators for whether the buyer: flips the home within a year; has an address within the same county as the property; has a non-U.S. address; is an institutional buyer; and purchases another home all-cash over our sample period. Together, these characteristics plausibly describe the seller's motivation to complete the transaction smoothly and the buyer's ability to identify such sellers.

The results in Table 5 imply a mortgage-cash premium between 10.0% and 12.9%. Notably, the buyer controls function similarly to the variable $Mortgaged_{i,j,t}$ from Table 4 in that they absorb price-relevant characteristics of mortgaged buyers, such as the propensity to overpay in out-of-town markets (e.g., Chinco and Mayer 2016. Relative to Table 4, however, we now also control for seller characteristics. These characteristics reduce bias from the possibility that, for example, all-cash buyers strategically make a large number of unfairly cheap offers with the intent of uncovering a motivated seller.

[9] We identify sellers and buyers in the ZTRAX dataset using unique names, as described in Appendix A. This method has the advantage of transparency, but it inevitably introduces measurement error that attenuates the estimated coefficients on the controls towards zero. To reduce the impact of such measurement error, we specify the controls as indicator variables.

5.2.2 Seller Cash Preference as an Instrumental Variable

Next, we instrument for $Mortgaged_{i,t}$ using the share of homes sold by the seller over our sample period that are to cash buyers, excluding the sale in question. Thus, this instrument is a " leave-one-out-mean", as commonly used in the labor and development literatures (e.g., Townsend 1994), and so it may be interpreted as the persistent component of a seller's preference for all-cash buyers. Denoting this instrument by $Cash\ Share_{s(i,t)}$, our principal identification assumption becomes

$$\mathbb{E}\left[\, Cash\ Share_{s(i,t)} \times \epsilon_{i,t} \,\middle|\, \alpha_i, \zeta_{z(i),t}, X_{i,t}\right] = 0, \tag{9}$$

where $s(i,t)$ denotes the seller of property i in month t.

Assumption (9) us to identify the mortgage-cash premium under substantially weaker conditions than its counterpart, assumption (5). In particular, we can allow mortgaged transactions to command a higher price for reasons apart from the method of financing, as long as these reasons do not covary with the seller's persistent preference for cash buyers ($Cash\ Share_{s(i,t)}$). This persistent preference differs from the measures of transitory preference in Table 5, in that it proxies for a structural characteristic of the seller as opposed to a temporary pressure that jointly affects the listing price and the seller's preference for cash (e.g., moving shock).[10]

Columns (1)-(2) of Table 6 substantiate the previous paragraph. We verify the first stage in column (1), which shows that sellers with a high value of $Cash\ Share_{s(i,t)}$ are less likely to sell to mortgaged buyers. The first stage is strong by the Stock and Yogo (2005) criteria. In column (2), we follow other papers in the literature (e.g., D'Acunto and Rossi 2020) and evaluate assumption (9) by reestimating our baseline regression equation (4) after including $Cash\ Share_{s(i,t)}$ as a control. We obtain a highly insignificant coefficient on $Cash\ Share_{s(i,t)}$. This finding supports the instrument's validity by showing that it has no effect on a transaction's sales price once controlling for the method of financing.

[10]This instrument also avoids bias from property condition, which we take up in the next subsection. Intuitively, a property in poor condition may attract all-cash buyers, but the property's seller was determined in its previous sale when its condition was plausibly different.

Columns (3)-(4) of Table 6 summarize the second stage results. We estimate a mortgage-cash premium of 15.6% in column (3). In column (4), we control for the seller characteristics from Table 5. This specification gives a similar estimate of 13.9%. In fact, both estimates are so similar to their counterpart in the baseline Table 2 that the Durbin-Wu-Hausman (DWH) test fails to reject the null hypothesis that $Mortgaged_{i,t}$ satisfies the repeat-sales-hedonic identification assumption (5).

Combining Tables 5 and 6, we obtain a mortgage-cash premium between 10% and 15% after firmly shutting down any variation in $Mortgaged_{i,t}$ that could covary with the seller's reservation value. This finding strongly supports the internal validity of an 11% premium.

5.3 Property Condition

The final class of internal validity violations stems from public information about the property's condition, which buyers and sellers observe but we as econometricians do not. Importantly, such violations cannot occur through time-invariant features of the property, as these are absorbed by the property fixed effect α_i, nor through time-varying features of the local market, which the zip code-month fixed effect $\zeta_{z(i),t}$ absorbs. Rather, we obtain upwardly biased estimates only if all-cash buyers tend to purchase properties temporarily in poor condition. The offer-level research design from Section 5.1 already addresses many such concerns. We build on that research design through three additional exercises, the details of which are in Appendix B.

First, we use data from the California Association of Realtors (CAR) to control for a property's condition through two important characteristics of the listing: the number of days on the market; and the listing price. Thus, the CAR dataset helps address specific concerns about property condition, but we cannot use it for our main analysis as it only covers a single state and a single year, 2019. This exercise yields an estimated premium of 14.3%, shown in Appendix Table A2.

Second, we estimate the premium through propensity score matching, a nonparametric approach that reduces the scope for bias by restricting the comparison to highly similar properties within the same zip code and year. Our implementation follows Harding, Rosenblatt and Yao (2012), who use this approach to estimate the foreclosure discount. We obtain an estimated premium of 16.9%, shown in Appendix Table A3.

Third, we implement the semi-structural estimator of Bajari et al. (2012), which allows us to recover the mortgage-cash premium even if the method of financing covaries with property condition. This approach explicitly models property condition as a Markov process, leading to a moment condition that expresses the sales price as a function of both the contemporaneous and lagged method of financing. Appendix A4 reports the estimated premium of 14.9%.

Together, these three exercises suggest that unobserved property condition does not upwardly bias the estimated premium of around 11%, supporting its internal validity.

5.4 External Validity

By prioritizing internal validity, we must necessarily restrict the variation used to identify the mortgage-cash premium (e.g., repeat sales), which raises questions of external validity that we assess in Appendix B.

First, in Appendix Table A1, we estimate premiums of: 18.6%, using the 11,367,195 transactions in the ZTRAX universe; 12.6%, using the 3,911,805 transactions in the subset of the ZTRAX universe that occur in the same zip code-by-month bins as in our main analysis; and 16.1%, using the 2,254,389 transactions in our main analysis without the inclusion of a property fixed effect.

Second, also in Appendix Table A1, we estimate a premium of 12.2% when applying our repeat-sales-hedonic approach to a nationally-representative, transaction-level dataset from CoreLogic. This regression features an R-squared of 95% despite a very large sample of 62,547,998 transactions, which strongly supports the estimate's internal validity. Moreover, that we obtain almost the same estimate from an entirely different data vendor strongly supports the external validity of our baseline, Zillow-based results.

Third, in Appendix Table A5, we weight transactions by their inverse probability of appearing in the baseline sample from Table 2, based on observed characteristics (e.g., Solon, Haider and Wooldridge 2015). This exercise yields an estimated premium of 10.3%.

Lastly, in Appendix Table A6, we show that our results agree with contemporaneous papers studying the price of cash-financed purchases, after adjusting for differences in sample (e.g., Buchak et al. 2020; Han and Hong 2020). Specific details are in Appendix B.4.

Collectively, these findings support the external validity of an 11% premium.

6 Calibrated Framework

A positive mortgage-cash premium does not contradict the original Modigliani and Miller (1958) logic as long as traditional transaction frictions can explain its magnitude. We evaluate this possibility by calibrating a framework of rational home sellers facing realistic transaction frictions. First, we emphasize the key intuition through a simple setup (6.1), which we then extend (6.2).

6.1 Basic Framework

Consider the seller of property i in month t. She has received offers from two buyers interested in purchasing i. The first buyer would purchase i exclusively with cash, and the second buyer would purchase it using mortgage financing. If the seller accepts the all-cash offer, she receives utility

$$u_{i,t}^{C} = \frac{\left(P_{i,t}^{C}\right)^{1-\gamma}}{1-\gamma}, \tag{10}$$

where, as in Section 4.1, $P_{i,t}^{C}$ is the sales price in the state of the world that the purchase is cash-financed; and $\gamma > 1$ is the seller's coefficient of relative risk aversion.

A seller who instead accepts a mortgaged offer must wait one month before learning the transaction's outcome, which approximates the average mortgage closing period (Ellie Mae 2012). The transaction can either fail, which occurs with probability q, or it can succeed, which occurs with probability $1-q$. In the latter case, the seller's utility is

$$u_{i,t}^{M}=e^{-\rho}\frac{\left(P_{i,t}^{M}\right)^{1-\gamma}}{1-\gamma},\tag{11}$$

where $P_{i,t}^{M}$ is the sales price in the state of the world that the purchase is mortgage-financed; and ρ is the rate at which the seller discounts utility over a one-month time horizon, with r denoting the equivalent rate at which consumption in one month is discounted.[11] Note that this setup abstracts from two realities: cash-financed purchases fail, and they do not close immediately. Our analytic results are approximately correct after reinterpreting q as the difference in the probability of failure between mortgage and cash-financed purchases and, similarly, reinterpreting ρ as the discount rate over the difference in closing periods. We shall account for these simplifications in our subsequent calibration.

In the case of failure, the seller relists the property at price

$$P_{i,t+1}^{R}\equiv P_{i,t}^{C}e^{-\kappa},\tag{12}$$

where κ equals the log price cut relative to the all-cash offer. The choice of κ determines the probability of attracting an offer in month $t+1$ at the relisted price, as we soon endogenize in Section 6.2. For now, suppose that the seller has a deadline, and, consequently, she must successfully sell her property by the end of month $t+1$. While extreme, this urgency to sell simplifies the analysis and provides an upper bound on the theoretical mortgage-cash premium. We suppose that κ is sufficiently large to attract an all-cash offer at $t+1$, which, given her time constraints, the seller then immediately accepts.

In equilibrium, the seller is indifferent between accepting an all-cash offer and a mortgaged one. Explicitly,

$$\frac{\left(P_{i,t}^{C}\right)^{1-\gamma}}{1-\gamma}=e^{-\rho}\left[(1-q)\frac{\left(P_{i,t}^{M}\right)^{1-\gamma}}{1-\gamma}+q\frac{\left(P_{i,t+1}^{R}\right)^{1-\gamma}}{1-\gamma}\right]\tag{13}$$

[11]Unlike ρ, this consumption-equivalent discount rate has the attractive feature that higher values of r lower the present value of future utility for any value of $\gamma\neq 1$. Appendix C shows that $e^{\rho}=e^{r(1-\gamma)}-1$.

Intuitively, the seller can accept the all-cash offer and receive the price $P_{i,t}^C$ immediately, which yields the utility shown on the left side of the equation. Alternatively, she can accept the mortgaged offer, wait an additional month, and subsequently learn whether the transaction has succeeded, in which case she receives the price $P_{i,t}^M$, or whether it has failed, in which case she receives the reduced price $P_{i,t+1}^R$.

Equation (13) assumes that sellers price the mortgage-cash premium, not buyers. Otherwise, the equilibrium premium would adjust to make buyers indifferent between mortgage and cash financing. A seller-based pricing function has three advantages over a buyer-based pricing function. First, First, the multimodal distribution of LTV ratios in Figure 1 suggests that most buyers either face binding borrowing constraints or do not borrow at all, which is inconsistent with them being indifferent between methods of financing. Second, most mortgaged buyers commit to their method of financing prior to searching for homes by obtaining loan pre-approval. Third, the average seller receives 2.6 offers (NAR 2018, 2020), which is consistent with sellers deciding between methods of financing as in equation (13). Although sellers may not always receive both cash and mortgage-financed offers, it seems reasonable to suppose that they are kept to their reservation utility by the possibility that a second offer under the alternative financing method will arrive.

The mortgage-cash premium, μ, adjusts such that equation (13) holds in equilibrium. Recall from equation (3) that the mortgage-cash premium links $P_{i,t}^M$ and $P_{i,t}^C$ according to

$$P_{i,t}^M = P_{i,t}^C e^{\mu}. \tag{14}$$

Equation (13) then gives the following expression for the theoretical mortgage-cash premium,

$$\mu^* = \frac{1}{\gamma - 1}\left[q\left(e^{\kappa(\gamma-1)} - 1\right) - e^{r(1-\gamma)} + 1\right], \tag{15}$$

as described in Appendix C.

Intuitively, sellers require mortgaged buyers to pay a larger premium μ when mortgaged transactions are more likely to fail (q). Similarly, when sellers expect to receive a substantially reduced price after failure, then they require mortgaged buyers to pay more (κ). Sellers must also be compensated for the mere fact that it takes longer for mortgaged transactions to close because they are impatient (r). Lastly, the mortgage-cash premium is greater when sellers are more risk-averse (γ).

6.2 Extended Framework

Our subsequent calibration and survey design rely on an extended framework that incorporates: seller debt overhang, simultaneous down payments, real estate agent fees, mortgage-specific closing costs, home search costs and a micro-foundation for the price cut after transaction failure.

First, suppose that the seller must pay her remaining mortgage balance, $L_{i,t}$, regardless of which offer she accepts. Second, she must pay $D_{i,t}$ to cover any remaining expenses, including a down payment on another property, moving costs, or any other cost apart from real estate agent commissions, which equal a share $1 - e^{-f}$ of the sales price. For ease of language, we shall simply refer to $D_{i,t}$ as the seller's "simultaneous down payment". Third, if the seller accepts the mortgaged offer, she must pay mortgage-specific closing costs equal to a share $1 - e^{-\xi}$ of the contractual sales price. We interpret these costs as the result of unmodeled bargaining between buyers and sellers.

Lastly, in the event of failure, suppose that sellers must cancel any contract they have to buy another home concurrently. Then, they search for a new home with their desired characteristics. The cost of search equals a share $1 - e^{\sigma}$ of the relisted price, where the parameter σ typically has the interpretation of time and effort spent (e.g., Guren 2018; Piazzesi, Schneider and Stroebel 2020). We do not interpret σ as the loss of an earnest money deposit, which seems reasonable given that the home seller would retain a similar deposit made by the mortgaged buyer who triggers the failure.

These extensions imply the following theoretical mortgage-cash premium,

$$\mu^* = \xi + \frac{1}{\gamma - 1}\left[q\left(\left[\frac{1-\ell-\delta}{e^{-\kappa-\sigma}-\ell-\delta}\right]^{\gamma-1} - 1\right) - e^{r(1-\gamma)} + 1\right], \tag{16}$$

where ℓ and δ are similar to the current loan-to-value ratio and the simultaneous down payment ratio, respectively,

$$\ell \equiv \frac{L_{i,t}}{P^C_{i,t} e^{-f}}, \quad \delta \equiv \frac{D_{i,t}}{P^C_{i,t} e^{-f}}. \tag{17}$$

Equation (16) implies that sellers require a higher premium when their own mortgage debt overhang is larger (ℓ), when they have a simultaneous down payment (δ), or when they must expend substantial effort searching for a replacement home (σ).

Appendix C provides a micro-foundation for the price cut after failure, κ, which recognizes that sellers cannot guarantee the arrival of an offer in month $t+1$. Specifically, we consider an exogenous offer arrival probability, and we suppose the seller continues to cut her list price until, eventually, an offer does arrive.

6.3 Calibration

We calibrate μ using equation (16), which nests the more basic expression in equation (15). Our set of parameters fall into two categories: the preference parameters r and γ; and the mortgage market parameters q, κ, ξ, ℓ, and δ. We choose relatively traditional preference parameters of $r = 3\%$, on an annualized basis, and $\gamma = 5$. The values of the remaining externally calibrated parameters come from industry and academic sources described below, with details in Appendix C.

We first parameterize $q = 6\%$ based on the average annual share of purchase transactions that were terminated over 2015-2021 according to the National Association of Realtors (NAR) dataset referenced in Section 2. Of this share, 54% of failures stemmed from "issues obtaining financing, appraisal issues, or home inspection issues", all of which pertain directly to the standard underwriting, appraisal, and inspection conditions required by most mortgage lenders. Thus, $q = 6\%$ reasonably approximates the precise theoretical object, which is the difference in conditional failure rates between mortgage and cash-financed purchases. Appendix C.4 shows how we obtain similar results when correcting for this simplification.

As an alternative, we also parameterize $q = 8.2\%$ based on the mortgage application denial rate over our sample period. To match our framework, we calculate the denial rate among pre-approved, first-lien applications for the purchase of an owner-occupied, one-to-four family home. Parameterizing q in this fashion provides an upper bound, since not all mortgage denials necessitate that the buyer terminate the purchase.

We parameterize $\kappa = 5.2\%$ based on the forced sale discount estimated by Campbell, Giglio and Pathak (2011). This estimate comes from comparing ordinary transactions with those in which the previous owner experiences sudden illness or death. Like the seller from our framework, the sellers in such transactions have a strong urgency to complete the sale. We intentionally omit the Campbell, Giglio and Pathak (2011) estimate of forced sales on foreclosed properties, since this discount captures the effects of both urgency, our desired channel, and quality. Instead, we also parameterize $\kappa = 5.7\%$ based on the foreclosure discount estimated in Figure 3 of Harding, Rosenblatt and Yao (2012), who isolate the role of urgency through an intensive matching procedure. Lastly, we parameterize $\kappa = 6.0\%$ according to the micro-foundation described in Appendix C.

The remaining objects to parameterize are ξ, ℓ, δ, and σ. We describe the choice of these parameter values in Appendix C. Lastly, we parameterize $f = 6\%$ to match the customary real estate agent fee paid by sellers. Table 7 summarizes all parameter values.

6.4 Theoretical Results

Table 8 summarizes the calibrated mortgage-cash premium. Columns (1)-(5) report a calibrated premium between 0.7% and 0.9% under various parameterizations of the basic framework described in Section 6.1. In columns (6)-(8), we report the results from applying our baseline parameterization to successively more rich versions of the extended framework in Section 6.2. Collectively, these additional features imply a calibrated premium of 3.0%.

A calibrated premium of 3.0% lies well below the most conservatively estimated premium of 8.5% shown in Table 3. We conclude this section by first assessing the scope for miscalibration and then by examining the time-series and cross-sectional co-movement between the calibrated and estimated premiums.

6.4.1 Robustness to Miscalibration

Households may not solve the model calibrated in Table 8 because we have either miscalibrated it or because we have omitted an important nontraditional channel. Section 7 studies the latter possibility. Here, we consider the scope for miscalibration by plotting the values of q and κ implied by a given value of the mortgage-cash premium. This exercise informs whether the results in Table 8 hold generically or only in a knife-edge parameterization.

Figure 3 shows how a premium of $\mu = 11\%$ and a price cut of $\kappa = 6\%$ implies a probability of failure $q = 36\%$, or six times the long-run average reported in Table 7. Matching this long-run average under an 11% premium requires a price cut of 18%, or three times the forced sale discount accepted by the literature. Even a premium of 8% requires values of q and κ that lie closer to their values during the 2008 crisis, shown by the red open circle. Appendix Figure A1 shows similar findings in terms of the home search cost σ and coefficient of relative risk aversion γ, with, for example, risk aversion of $\gamma = 23$ required to explain an 11% premium. Thus, traditional modelling decisions predict a relatively small mortgage-cash premium unless we accept extreme parameter values.

6.4.2 Time Series of the Estimated and Calibrated Premiums

Next, we recalibrate the mortgage-cash premium for each of the time periods shown in columns (2)-(5) of Table 2. Then, we plot the calibrated and estimated premiums together in Figure 2. The calibrated and estimated premiums co-move over time. Both values peak over 2005-2010, which, as described in Appendix C, stems from elevated debt overhang (ℓ), price cuts (κ), and transaction risk (q). However, the gap between the estimated and calibrated premiums remains stable at between 8% and 10%. Appendix Figure A2 shows how the calibrated and estimated premiums also co-move over the amount of leverage in a transaction, but a similarly stable gap obtains.

6.4.3 Estimated Premium by the Degree of Transaction Friction

Lastly, Table 9 reestimates equation (4) after interacting $Mortgaged_{i,t}$ with empirical proxies for the parameters ℓ, δ, κ, and q, all of which govern degree of friction in a transaction.[12] Columns (1) and (2) estimate a higher premium when the seller's existing loan-to-value ratio (ℓ) and down payment on the next home she buys (δ) exceed the sample average, respectively. Column (3) estimates a higher premium when the number of transactions in the zip code-by-month bin lies below average, a proxy for illiquidity and, thus, a higher value of κ. Lastly, column (4) estimates a higher premium when the mortgage-application denial rate in the surrounding zip code-by-month bin exceeds the long-run average denial rate according to HMDA. As with Figure 2, these findings show that the mortgage-cash premium co-moves with the degree of friction. However, the estimated coefficient on $Mortgaged_{i,t}$ throughout Table 9 suggests that even mortgaged purchases with relatively little friction still command a premium between 8% and 10%.

Collectively, the empirical mortgage-cash premium estimated in Section 4 exceeds the premium predicted by a traditional model with frictions by a factor of four or, taking the most conservative estimate from Section 5, a factor of three. This latter comparison obtains after an exhaustive investigation of selection bias. Therefore, we turn our attention away from the magnitude of the empirical premium and ask whether a nontraditional model can provide a more realistic magnitude of the theoretical premium.

7 Nontraditional Channels

Broadly, nontraditional explanations of the mortgage-cash premium puzzle can fall into two categories (e.g., DellaVigna 2009). First, home sellers may solve a model similar to that from Section 6, but their prior beliefs about mortgage transaction risk differ from the parameterizations in Table 7. Second, nontraditional preferences or optimization may lead sellers to solve an entirely different model, which, as Barseghyan et al. (2013) note, leads

[12] As in Table 5, we specify these empirical proxies as indicator variables to reduce the impact of measurement error.

them to act as if they were facing a distorted probability of failure. We use a survey to assess the relative merit of several nontraditional explanations. As a side benefit, we can further assess the validity of our baseline results by estimating the premium with experimental data. We describe the survey's design (7.1), summarize the data (7.2), and quantify the role of heterogeneous priors (7.3) and probability distortions (7.4).

7.1 Survey Design

Following a number of recent economics papers summarized by Lian, Ma and Wang (2018), we administer the survey online to participants who have been pre-screened, compensated for their participation, and recruited through a mainstream crowdsourcing platform. This approach is logistically efficient and can produce results of comparable quality to those obtained in laboratory experiments (e.g., Casler, Bickel and Hackett 2013). We specifically partner with the crowdsourcing platform Prolific, a competitor to Amazon's commonly used MTurk platform. Relative to MTurk, Prolific allows us to exclusively recruit U.S. homeowners, which is critical for comparability with our baseline results. We administered the survey in two non-longitudinal waves with a similar question structure.

At a high level, the survey consists of a thought experiment in which we ask respondents to imagine that they are selling their current home. We specify a particular list price, chosen to match the typical sales price in the respondent's neighborhood. The conditions of the sale are similar to those in column (8) of Table 8. Specifically, we tell respondents that they have an outstanding mortgage balance equal to 30% of this list price (i.e., $\ell e^f = 30\%$). In addition, respondents are under contract to purchase a new home, and the associated down payment equals 15% of the current home's list price (i.e., $\delta e^f = 15\%$). To match our framework, this down payment must be made within six weeks, which also equals the mortgage closing period. Importantly, we state all quantities in both percent and dollar terms.[13]

[13] Here is an example transcript, divided over several screens: *"Imagine, also, that you have now listed your current home at $300,000. The remaining mortgage balance that you owe on it is $90,000. So, what you owe on the home is 30% of what you have listed it at. To make the down payment on the new home, you will need to finalize the sale of your current home, pay off the balance, and then use the money that's left over. The down payment is $45,000. That is 15% of the price at which you have listed your current home."*

After describing the conditions of the sale, we ask respondents to imagine that they have received offers from two buyers. The first buyer would pay exclusively with cash, and the second buyer would pay using mortgage financing. We specify that the all-cash transaction has "almost no risk" of failing and will close "any time within two weeks". The mortgaged transaction would close in six weeks, thus maintaining the one-month lag from our framework. In addition, "there is a chance that the mortgaged buyer will not be able to secure money from their lender", in which case the respondent "will need to relist your home in six weeks." In the survey's second wave, a random 50% subset of participants are told that mortgaged transactions will fail 7% of the time, consistent with contemporaneous market data (NAR 2021a, 2021b), and that they would need to impose a 6% price cut to attract another offer, again expressed in dollars and percent.

The remainder of the survey consists of four blocks of questions. In the first block, we tell respondents that both the all-cash and the mortgaged buyer offer to pay their list price, $\$B$, and then we ask them which offer they would prefer. Conditional on preferring the all-cash offer, we ask respondents to select their strongest reason for doing so from a list of three predefined options and a fourth free-response option.

The second block of questions elicits the respondent's mortgage-cash premium through a sequence of pairwise comparisons (i.e., multiple price list), following standard practice. We ask the respondent which offer she would prefer at gradually increasing spreads between the mortgaged and the all-cash offer price. Each question takes the form: *"Suppose the Mortgaged Buyer offers to pay* $\$[(1+\tilde{\mu})\times B]$*. That is* $[100\times\tilde{\mu}]\%$ *more than the Cash Buyer. Which offer would you accept now?"* The spread $\tilde{\mu}$ grows in increments of 4 pps until reaching 28%. Most respondents switch from preferring the all-cash offer to the mortgaged offer once $\tilde{\mu}$ exceeds some threshold. Define the mortgage-cash premium for respondent k, denoted $Premium_k$, as the midpoint between the minimum value of $\tilde{\mu}$ at which she prefers the mortgaged offer and the maximum value of $\tilde{\mu}$ at which she prefers the all-cash offer.[14] For simplicity, we call $Premium_k$ the "experimental premium".

[14]We assign a missing value to participants who exhibit multiple switch points (e.g., Bernheim and Sprenger 2020). In the survey's first wave, we increased $\tilde{\mu}$ in increments of 5 pps. We reduced the step size to improve accuracy, but Table 10 shows how this does not affect the results.

In the third block, we elicit respondents' prior beliefs about mortgaged transaction failure (q) and the cost of such failure (κ) for the subset who are not already given these values. Respondents select a value for q on a sliding scale with an upper bound of 30%. The use of a sliding scale does not lead to bunching around the scale's midpoint, since 42% of respondents select a value less than 10% or greater than 20%. Likewise, Respondents select the price level at which they would relist their home after failure, corresponding to $Be^{-\kappa}$, and the equivalent percent price cut.

Finally, we elicit the respondent's numeracy following the method of Lipkus, Samsa and Rimer (2001) as described in Appendix D.2, and we collect information about the respondent's annual household income, age, state of residence, education, and risk aversion (e.g., Fuster and Zafar 2021). In the first wave, we also collected information about the number of homes sold by the respondent.

Our resulting dataset contains information on 938 U.S. homeowners from the survey's first wave and 1,105 from the second wave. We administered the two waves in April 2021 and November 2021, respectively. This period saw extraordinarily favorable conditions for sellers, with strong demand, limited supply, and low interest rates (Gascon and Haas 2020). These conditions make it less costly to accept a risky mortgage-financed offer (e.g., low κ, q), and so the mortgage-cash premium elicited through this survey is likely conservative.

7.2 Experimental Mortgage-Cash Premium

Table 10 reports an average elicited premium 10.4%, shown in the table's first row. The average premium is consistent across the survey's first wave (10.2%) and second wave (10.5%), where the latter statistic comes from averaging columns (3)-(5). These values lie close to our baseline estimate of 11.7% from Table 2, and they fall well within the range of additional estimates shown in Table 3. That we obtain similar results from both experimental and observational data provides strong support for our baseline finding.

Turning to the background information summarized in the bottom panel, respondents represent U.S. homeowners relatively well. If anything, they have higher levels of education and income. For example, the share of respondents with a bachelor's degree (69.4%) exceeds the average share among U.S. homeowners (40.1%), per the 2019 American Housing Survey. Geographically, Appendix Figure A3 shows how the survey covers all U.S. states except Wyoming in similar proportion to state population.

We briefly highlight two additional observations to which we return in detail below. First, column (1) shows how respondents not facing a given distribution of outcomes hold pessimistic priors regarding the probability of failure (13.4%), relative to historical data.[15] Second, column (5) shows how respondents require a higher premium (12%) when they do not face a given distribution but nevertheless hold priors that lie within 1 pps of it. Together, these observations suggest that a combination of heterogeneous priors and uncertainty aversion can explain the average mortgage-cash premium.

7.3 Heterogeneous Priors

Maintaining the traditional preferences from Section 6, suppose sellers hold heterogeneous priors about the probability of transaction failure (q) and, similarly, vary in the price cut they would impose after failure (κ). We evaluate the importance of such heterogeneity by using our survey dataset to plot the relationship between respondents' elicited premium and the premium implied by their beliefs, shown in Figure 4.

As a benchmark, the blue open circle shows how respondents who are given values of q and κ consistent with historical data require a premium that lies 8.9 pps above the premium implied by this parameterization. This finding replicates the puzzle documented in our observational datasets, marked by the red triangle.

By contrast, respondents who do not face a given distribution hold priors that imply a premium between 0.5% and 15%. The puzzle's magnitude shrinks as respondents become more pessimistic, shown by the declining distance between the blue solid circles and the

[15] The term "relatively pessimistic" is appropriate both in terms of historical failure probabilities shown in Table 7 and the contemporaneous failure probability of 7% (NAR 2021a, 2021b). In particular, the average prior failure probability of 13.4% doubles its historical counterpart in Table 7. Note that the price cut that the average respondent would impose after failure of 4.5% lies close to the benchmark forced sale discount of 5.2% shown in Table 7, and so we use "pessimism" in reference to the probability of failure.

45-degree line. Since the average respondent is relatively pessimistic, the average gap between the elicited and implied premium is smaller than among respondents facing a given distribution.

Quantitatively, columns (1) and (2) of Table 11 show how the average premium gap falls from 8.9% to 7.2% when moving from the sample of respondents who are given values of q and κ to the sample in which respondents rely on their own priors.[16] Thus, heterogeneous priors can account for 19% (1.7 pps) of the puzzle.

7.4 Probability Distortions

Next, suppose sellers make decisions using distorted values of their own prior probabilities of failure: the probability they have in mind differs from the probability they use in optimization. We evaluate the importance of four specific distortions, all of which require introducing nontraditional preferences: uncertainty aversion, disappointment aversion, loss aversion, and probability weighting.

In the following exercise, we make the minimal adjustments necessary to incorporate each distortion into our framework. Then, using the literature's accepted parameter values, we quantify the distortion's ability to explain the premium gap. Although we take a structural approach, our goal is not to draw precise numerical conclusions but, rather, to provide a ranking of candidate distortions. We defer details on the derivation and calibration of each distortion to Appendix C.

7.4.1 Uncertainty Aversion

We incorporate uncertainty aversion following Hansen and Sargent (2011). Applying the logic to our setting, each seller has a prior probability of failure but does not know the true probability. Accordingly, she evaluates mortgaged offers under the "worst-case"

[16]The latter statistic is calculated holding κ fixed at 6% to emphasize the role of beliefs, and we obtain an average gap of 6.3% when relaxing this restriction. All results presented in this section weight respondents by time spent on the survey to reduce noise from inattentive respondents, but the results do not differ when not weighting.

probability of failure. She obtains this worst-case probability by choosing the value of q that minimizes her expected utility, subject to a cost defined by the entropy of q relative to her prior. The parameter θ governs the importance of this cost: when $\theta = 0$, the seller exhibits no uncertainty aversion and imposes a high cost on probabilities that depart from her prior; as $\theta \to \infty$, the seller exhibits such strong uncertainty aversion that she disregards her prior, evaluating mortgaged transactions as though they fail with probability one.

Appendix C shows how the resulting worst-case probability equals

$$q^U = q \left[q + (1-q)\, e^{-\theta\left(\frac{u^M - u^R}{(P^C)^{1-\gamma}}\right)} \right]^{-1}. \tag{18}$$

In words, q^U describes a probability distortion that increases in the seller's aversion to uncertainty (θ) and in the normalized utility loss from the worst-case outcome ($\frac{u^M - u^R}{(P^C)^{1-\gamma}}$). We calculate q^U under the parameterization $\theta = 0.33$ and, to avoid mechanical correlation, using the average utility loss across survey respondents.[17] Notably, parameterizing $\theta \leq 1$ would be considered conservative by the literature's standards, which commonly features much greater uncertainty aversion (e.g., Maenhout 2004; Barnett, Buchak and Yannelis 2021).

Columns (3)-(4) of Table 11 show how uncertainty aversion can reduce the average premium gap to between 0.7% and 1.6%. With modestly higher aversion of $\theta = 0.36$, the premium gap would actually equal zero. To understand why uncertainty aversion performs well, the blue solid circles in Figure 5 plot the failure probability implied by respondents' elicited premium against their prior probabilities. This relationship demonstrates substantial curvature, which the functional form in equation (18) fits well, per the blue dashed line. Intuitively, the convex entropy cost function that microfounds equation (18) leads sellers to make decisions based on a similar worst-case probability, regardless of their prior.

[17] This parameterization follows from the calibration strategy developed by Anderson, Hansen and Sargent (2003). Essentially, this method ensures that sellers will never choose a worst-case probability that is rejectable at some reasonable threshold (e.g., 5%), based on available data. We obtain $\theta = 0.33$ when defining "available data" using the average number of sales among survey respondents and $\theta = 0.30$ when using census tract-by-year data on mortgage application denial rates.

7.4.2 Disappointment Aversion

We incorporate disappointment aversion following Gul (1991). Accordingly, home sellers experience disappointment when a mortgaged transaction fails because failure results in lower utility than the transaction's certainty equivalent.[18] Since disappointment only occurs in failure, disappointment aversion functions similarly to a distortion of the failure probability. In particular, a disappointment averse seller evaluates mortgaged offers according to

$$q^D = \frac{q\left[1+\beta\right]}{1+\beta q}, \tag{19}$$

where $\beta \geq 0$ governs the additional utility loss from disappointment. Typically, β ranges from 1 to 2 (e.g., Epstein and Zin 1991; Ang, Bekaert and Liu 2005;Bonomo et al. 2011).

Columns (5)-(6) of Table 11 show how disappointment aversion can reduce the average premium gap to between 3.9% and 4.5%. Fully explaining the gap, though, would require aversion of $\beta = 5.2$, around four times the literature's convention. In particular, Figure 5 shows how disappointment aversion cannot explain the behavior of respondents with a low prior probability of failure, unlike uncertainty aversion.

7.4.3 Loss Aversion

We incorporate loss aversion following Kőszegi and Rabin (2006). Like disappointment aversion, loss aversion induces reference-dependent preferences. Unlike disappointment aversion, however, we must now select an appropriate reference point. A long tradition in the real estate literature has focused on the role of the purchase price as a reference point (e.g., Genesove and Mayer 2001). Yet, unless the purchase price lies in the narrow range between the relisted price after failure and the price of an all-cash offer, its role as a reference point cannot explain the mortgage-cash premium.

Instead, we set the list price as the reference point. This accords with how 90% of survey respondents concerned with meeting a target cite the list price as that target, per Appendix

[18] Equation (13) implies that the all-cash offer constitutes the certainty equivalent of the mortgaged transaction, and so our framework cannot distinguish between disappointment aversion and regret aversion.

Table A7. In the survey's first wave and in half of the second wave, the all-cash offer equals or exceeds the list price. Under this ordering, a loss averse seller requires mortgaged buyers to pay a premium as compensation for the risk of not receiving the list price.

Applying the Kőszegi and Rabin (2006) framework to our setting, sellers value an offer in terms of the traditional expected utility shown in Section 6 plus the expected gain or loss in utility relative to receiving their list price. The parameter $\eta \geq 0$ governs the relative importance of gain-loss utility, and the parameter $\lambda \geq 1$ governs the degree of disutility from loss. As with disappointment aversion, loss aversion leads sellers to act according to a distorted probability,

$$q^L = \frac{q\left[1+\Lambda\right]}{1+\Lambda q}, \tag{20}$$

with $\Lambda \equiv \frac{\eta}{1+\eta}(\lambda - 1)$. As described in Appendix C, the literature typically parameterizes $\Lambda \leq 1.25$ (e.g., Tversky and Kahneman 1992; Barberis, Jin and Wang 2021).

The similarity between equations (19) and (20) implies that loss aversion functions similarly to disappointment aversion. Indeed, column (7) of Table 11 shows how loss aversion reduces the average premium gap to the same 4.5% as its peer in column (5). However, neither form of reference dependence can fully explain the premium gap without a parameterization that is at least four times as aggressive as that typically found in the literature.

7.4.4 Probability Weighting

Probability weighting refers to the tendency to perceive small probabilities as larger than they actually are when making decisions and, conversely, to perceive large probabilities as smaller. The literature has developed weighting functions that describe this tendency. We follow Tversky and Kahneman (1992), who propose the following functional form

$$q^W = \frac{q^\alpha}{\left[q^\alpha + (1-q)^\alpha\right]^{\frac{1}{\alpha}}}. \tag{21}$$

Many papers have estimated α, and the estimates commonly fall between 0.65 and 0.75 as summarized by Booij, Van Praag and Van De Kuilen (2010).

Probability weighting can only reduce the average premium gap to 5.7%, as shown in column (9) of Table 11. Moreover, there does not exist a value of α under which the premium gap equals zero. Intuitively, probability weighting compresses very small and very large probabilities closer to some central value. Therefore, probability weighting outperforms both disappointment and loss aversion when respondents have a prior failure probability less than 5%, as Figure 5 shows. However, since 78% of respondents have a prior of at least 10%, probability weighting struggles to reduce the average premium gap. Appendix C reaches a similar conclusion using the Prelec (1998) weighting function.

7.5 Discussion

Our structural analysis points to uncertainty aversion as the most relevant probability distortion. Three additional pieces of evidence support the contribution of uncertainty to the mortgage-cash premium. First, comparing columns (4) and (5) of Table 10, respondents who face uncertainty but hold priors similar to the given distribution require a 1.4 pps higher premium (12%) than those who are actually given this distribution (10.6%). This difference is consistent with an uncertainty premium.

Second, Table 12 shows how the mortgage-cash premium is lower among sellers with more experience, which we view as a proxy for less uncertainty. In columns (1)-(3), we regress a respondent's required premium on the number of homes she has sold, collected in the survey's first wave. The estimated coefficients imply that each additional home sold reduces the mortgage-cash premium by 0.6-1.0 pps. Notably, this result cannot confound any susceptibility of inexperienced sellers to sophisticated all-cash buyers, since this channel does not exist in our experiment. We address concerns of external validity by reestimating the repeat-sales-hedonic regression equation (4) after interacting $Mortgaged_{i,t}$ with the number of sales made by the seller as of month t in the baseline ZTRAX dataset. Interpreting the estimated coefficient on the interaction term in column (4), each sale reduces the mortgage-cash premium by 0.5 pps, similar to the effect estimated using experimental data.

Third, Appendix Table A7 shows how survey respondents facing an uncertain distribution are 13.2 pps (21%) more likely to cite the following as the most attractive aspect of an all-cash offer: "Even if the Mortgaged Buyer would never back out, the Cash Buyer would close more quickly and end the stressful process of selling my home." This statement does not strictly capture any particular channel, but its elevated popularity among respondents facing uncertainty suggests that knowing the distribution of transaction outcomes provides important stress relief.[19]

Collectively, our experimental evidence suggests that the combination of home sellers' uncertainty about mortgaged transaction outcomes and pessimistic beliefs about these outcomes leads them to require a large mortgage-cash premium.

8 Conclusion

We found that the capital structure of home purchases – mortgage versus cash – affects transaction value to an extent that cannot be explained by transaction frictions alone. Based on a variety of subsamples, estimators, datasets, and an experimental survey of U.S. homeowners, we consistently find that mortgage-financed home buyers must pay an 11% price premium relative to cash-financed buyers. By contrast, a traditional and realistically calibrated model of rational home sellers implies a premium of only 3%. Our survey points to uncertainty aversion and heterogeneous beliefs as the most promising explanation of the 8 pps discrepancy.

Our results have policy implications that derive from buy-side and sell-side perspectives on the mortgage-cash premium. From buyers' point of view, the mortgage-cash premium represents an additional cost of becoming a homeowner, since most first-time homebuyers rely on government-insured mortgages (e.g., Bai, Zhu and Goodman 2015). Consequently, a government interested in promoting homeownership must insure a large quantity of mortgage

[19] One might interpret this statement's popularity as evidence of present focus given the reference to "closing more quickly" (e.g., Laibson 1997; O'Donoghue and Rabin 1999). However, respondents randomly assigned to an all-cash offer closing "within two weeks" require a premium that is statistically indistinguishable from those whose all-cash offer closes "in four weeks", inconsistent with present focus. This finding may reflect how present focus concerns immediate consumption, whereas no home purchase can close immediately.

debt to accomplish this goal, relative to a frictionless counterfactual in which the mortgage-cash premium equals zero. From sellers' point of view, the mortgage-cash premium represents a large "cash discount". Therefore, a liquid housing market with more all-cash buyers may erode the value of real estate as a savings vehicle.

A lower mortgage-cash premium can come from easing transaction frictions or from reducing the nontraditional channels that amplify them. The former route has an outsized impact because of amplification. The latter route requires more research on how uncertainty varies across home sellers and why their prior beliefs differ from long-run data. We leave that question for future research.

References

Abadie, A. and Imbens, G. W.: 2006, Large Sample Properties of Matching Estimators for Average Treatment Effects, *Econometrica* .

Almeida, H. and Philippon, T.: 2007, The Risk-Adjusted Cost of Financial Distress, *The Journal of Finance* .

Andersen, S., Badarinza, C., Liu, L., Marx, J. and Ramadorai, T.: 2021, Reference Dependence in the Housing Market, *CEPR Discussion Paper* .

Anderson, E. W., Hansen, L. P. and Sargent, T. J.: 2003, A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection, *Journal of the European Economic Association* .

Ang, A., Bekaert, G. and Liu, J.: 2005, Why Stocks May Disappoint, *Journal of Financial Economics* .

Asabere, P. K., Huffman, F. E. and Mehdian, S.: 1992, The Price Effects of Cash versus Mortgage Transactions, *Real Estate Economics* .

Bai, B., Zhu, J. and Goodman, L.: 2015, A Closer Look at the Data on First-Time Homebuyers, *Urban Institute Working Paper* .

Bajari, P., Fruehwirth, J. C., Kim, K. I. and Timmins, C.: 2012, A Rational Expectations Approach to Hedonic Price Regressions with Time-Varying Unobserved Product Attributes: The Price of Pollution, *American Economic Review* .

Barberis, N., Jin, L. J. and Wang, B.: 2021, Prospect Theory and Stock Market Anomalies, *The Journal of Finance* .

Barnett, M., Buchak, G. and Yannelis, C.: 2021, Epidemic Responses Under Uncertainty, *NBER Working Paper* .

Barseghyan, L., Molinari, F., O'Donoghue, T. and Teitelbaum, J. C.: 2013, The Nature of Risk Preferences: Evidence from Insurance Choices, *American Economic Review* .

Barseghyan, L., Prince, J. and Teitelbaum, J. C.: 2011, Are Risk Preferences Stable across Contexts? Evidence from Insurance Data, *American Economic Review* .

Barwick, P. J. and Pathak, P. A.: 2015, The Costs of Free Entry: An Empirical Study of Real Estate Agents in Greater Boston, *RAND Journal of Economics* .

Bernheim, B. D. and Sprenger, C.: 2020, On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting, *Econometrica* .

Bleichrodt, H. and Pinto, H. L.: 2000, A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis, *Management Science* .

Bonomo, M., Garcia, R., Meddahi, N. and Tédongap, R.: 2011, Generalized Disappointment Aversion, Long-run Volatility Risk, and Asset Prices, *Review of Financial Studies* .

Booij, A. S., Van Praag, B. M. S. and Van De Kuilen, G.: 2010, A Parametric Analysis of Prospect Theory's Functionals for the General Population, *Theory and Decision* .

Buchak, G., Matvos, G., Piskorski, T. and Seru, A.: 2020, Why is Intermediating Houses so Difficult? Evidence from iBuyers, *NBER Working Paper* .

Cagetti, M., Hansen, L. P., Sargent, T. and Williams, N.: 2002, Robustness and Pricing with Uncertain Growth, *Review of Financial Studies* .

Campbell, J. Y., Giglio, S. and Pathak, P.: 2011, Forced Sales and House Prices, *American Economic Review* .

Caskey, J. A.: 2008, Information in Equity Markets with Ambiguity-Averse Investors, *Review of Financial Studies* .

Casler, K., Bickel, L. and Hackett, E.: 2013, Separate but Equal? A Comparison of Participants and Data Gathered via Amazon's MTurk, Social Media, and Face-to-Face Behavioral Testing, *Computers in Human Behavior* .

Chen, H.: 2010, Macroeconomic Conditions and the Puzzles of Credit Spreads and Capital Structure, *The Journal of Finance* .

Chen, L., Collin-Dufresne, P. and Goldstein, R. S.: 2009, On the Relation Between the Credit Spread Puzzle and the Equity Premium Puzzle, *Review of Financial Studies* .

Cheng, I.-H., Raina, S. and Xiong, W.: 2014, Wall Street and the Housing Bubble, *American Economic Review* .

Chinco, A. and Mayer, C.: 2016, Misinformed Speculators and Mispricing in the Housing Market, *Review of Financial Studies* .

D'Acunto, F. and Rossi, A.: 2020, Regressive Mortgage Credit Redistribution in the Post-Crisis Era, *Review of Financial Studies* .

DellaVigna, S.: 2009, Psychology and Economics: Evidence from the Field, *Journal of Economic Literature* .

Einav, L., Finkelstein, A., Pascu, I. and Cullen, M. R.: 2012, How General Are Risk Preferences? Choices under Uncertainty in Different Domains, *American Economic Review* .

Ellie Mae: 2012, Origination Insight Report, *Technical report*.

Epstein, L. G. and Zin, S. E.: 1991, The Independence Axiom and Asset Returns, *NBER Working Paper* .

Federal National Mortgage Association (Fannie Mae): 2019, Fannie Mae Selling Guide: Interested Party Contributions.

Federal Reserve Board: 2019, Financial Accounts of the United States: Mortgage Debt Outstanding, *Technical report*.

Foote, C. L., Loewenstein, L. and Willen, P. S.: 2020, Cross-Sectional Patterns of Mortgage Debt during the Housing Boom: Evidence and Implications, *Review of Economic Studies* .

Fuster, A. and Zafar, B.: 2021, The Sensitivity of Housing Demand to Financing Conditions: Evidence from a Survey, *American Economic Journal: Economic Policy* .

Gascon, C. S. and Haas, J.: 2020, The Impact of COVID-19 on the Residential Real Estate Market, *Federal Reserve Bank of St. Louis Regional Economist* .

Genesove, D. and Mayer, C.: 2001, Loss Aversion and Seller Behavior: Evidence from the Housing Market, *The Quarterly Journal of Economics* .

Gete, P. and Reher, M.: 2018, Mortgage Supply and Housing Rents, *Review of Financial Studies* .

Gete, P. and Reher, M.: 2021, Mortgage Securitization and Shadow Bank Lending, *Review of Financial Studies* .

Giglio, S., Maggiori, M. and Stroebel, J.: 2015, Very Long-Run Discount Rates, *The Quarterly Journal of Economics* .

Gilboa, I. and Schmeidler, D.: 1989, Maxmin Expected Utility with Non-Unique Prior, *Journal of Mathematical Economics* .

Gilbukh, S. and Goldsmith-Pinkham, P.: 2019, Heterogeneous Real Estate Agents and the Housing Cycle.

Glaeser, E. L. and Nathanson, C. G.: 2017, An Extrapolative Model of House Price Dynamics, *Journal of Financial Economics* .

Glaeser, E. L. and Sinai, T.: 2013, *Housing and the Financial Crisis*, The University of Chicago Press.

Greenwald, D.: 2018, The Mortgage Credit Channel of Macroeconomic Transmission.

Gul, F.: 1991, A Theory of Disappointment Aversion, *Econometrica* .

Guren, A. M.: 2018, House Price Momentum and Strategic Complementarity, *Journal of Political Economy* .

Han, L. and Hong, S.-H.: 2011, Testing Cost Inefficiency Under Free Entry in the Real Estate Brokerage Industry, *Journal of Business and Economic Statistics* .

Han, L. and Hong, S.-H.: 2020, Cash is King? Evidence from Housing Markets.

Hansen, L. P. and Sargent, T. J.: 2011, *Robustness*, Princeton University Press.

Hansen, L. P., Sargent, T. J., Turmuhambetova, G. A. and Williams, N.: 2002, Robustness and Uncertainty Aversion.

Harding, J. P., Rosenblatt, E. and Yao, V. W.: 2012, The Foreclosure Discount: Myth or Reality?, *Journal of Urban Economics* .

Hsieh, C.-T. and Moretti, E.: 2003, Can Free Entry Be Inefficient? Fixed Commissions and Social Waste in the Real Estate Industry, *Journal of Political Economy* .

Huang, J.-Z. and Huang, M.: 2012, How Much of the Corporate-Treasury Yield Spread Is Due to Credit Risk?, *Review of Asset Pricing Studies* .

Kaplan, G., Mitman, K. and Violante, G. L.: 2020, The Housing Boom and Bust: Model Meets Evidence, *Journal of Political Economy* .

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R. and Winter, N. J. G.: 2021, The Shape of and Solutions to the MTurk Quality Crisis, *Political Science Research and Methods* .

Kőszegi, B. and Rabin, M.: 2006, A Model of Reference-Dependent Preferences, *The Quarterly Journal of Economics* .

Laibson, D.: 1997, Golden Eggs and Hyperbolic Discounting, *The Quarterly Journal of Economics* .

Leuven, E. and Sianesi, B.: 2003, PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing.

Levitt, S. D. and Syverson, C.: 2008, Market Distortions when Agents are Better Informed: The Value of Information in Real Estate Transactions, *Review of Economics and Statistics* .

Lian, C., Ma, Y. and Wang, C.: 2018, Low Interest Rates and Risk Taking: Evidence from Individual Investment Decisions, *Review of Financial Studies* .

Lipkus, I. M., Samsa, G. and Rimer, B. K.: 2001, General Performance on a Numeracy Scale among Highly Educated Samples, *Medical Decision Making* .

Liu, H. and Palmer, C.: 2021, Are Stated Expectations Actual Beliefs? New Evidence for the Beliefs Channel of Investment Demand, *NBER Working Paper* .

Lusht, K. M. and Hansz, J. A.: 1994, Some Further Evidence on the Price of Mortgage Contingency Clause, *Journal of Real Estate Research* .

Maenhout, P. J.: 2004, Robust Portfolio Rules and Asset Pricing, *Review of Financial Studies* .

Mills, J., Molloy, R. and Zarutskie, R.: 2019, Large-Scale Buy-to-Rent Investors in the Single-Family Housing Market: The Emergence of a New Asset Class, *Real Estate Economics* .

Modigliani, F. and Miller, M. H.: 1958, The Cost of Capital, Corporation Finance and the Theory of Investment, *American Economic Review* .

National Association of Realtors: 2018, Realtors Confidence Index Survey.

National Association of Realtors: 2020, Realtors Confidence Index Survey.

O'Donoghue, T. and Rabin, M.: 1999, Doing It Now or Later, *American Economic Review* .

Peer, E., Brandimarte, L., Samat, S. and Acquisti, A.: 2017, Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research, *Journal of Experimental Social Psychology* .

Piazzesi, M. and Schneider, M.: 2009, Momentum Traders in the Housing Market: Survey Evidence and a Search Model, *American Economic Review* .

Piazzesi, M. and Schneider, M.: 2016, Housing and Macroeconomics, *Handbook of Macroeconomics.*

Piazzesi, M., Schneider, M. and Stroebel, J.: 2020, Segmented Housing Search, *American Economic Review* .

Prelec, D.: 1998, The Probability Weighting Function, *Econometrica* .

Rosenthal, S. S.: 2014, Are Private Markets and Filtering a Viable Source of Low-Income Housing? Estimates from a "Repeat Income" Model, *American Economic Review* .

Rubin, D. B.: 1974, Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, *Journal of Educational Psychology* .

Savage, L. J.: 1954, *The Foundations of Statistics*, John Wiley and Sons.

Seo, Y., Holmes, C. and Lee, M.: 2021, Examining the Cash-Only Price Discount and the Driving Forces of Cash-Only Transactions in the Housing Market, *Journal of Real Estate Finance and Economics* .

Shiller, R. J.: 2014, Speculative Asset Prices, *American Economic Review* .

Solon, G., Haider, S. J. and Wooldridge, J. M.: 2015, What Are We Weighting For?, *Journal of Human Resources* .

Stock, J. H. and Yogo, M.: 2005, Testing for Weak Instruments in Linear IV Regression, *Identification and Inference for Econometric Models*.

Townsend, R. M.: 1994, Risk and Insurance in Village India, *Econometrica* .

Trulia: 2009, Trulia Price Reduction Report: Listing Price Cuts on the Rise Nationally; Stabilization Seen in Some of the Hardest Hit Markets, *Technical report*.

Tversky, A. and Kahneman, D.: 1992, Advances in Prospect Theory: Cumulative Representation of Uncertainty, *Journal of Risk and Uncertainty* .

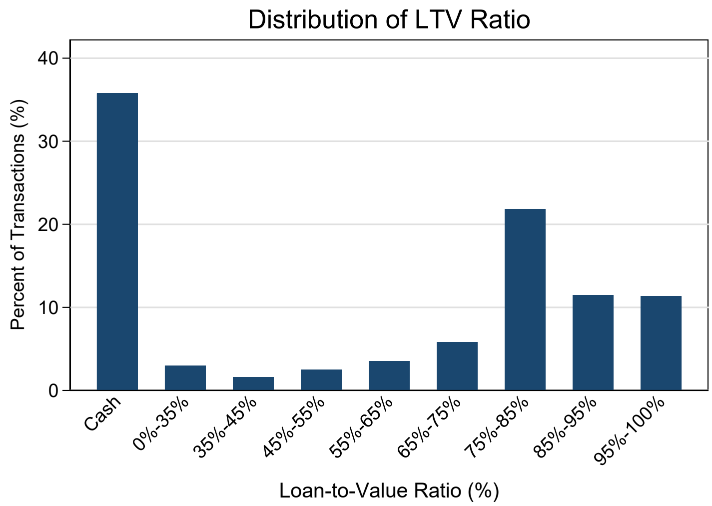
U.S. Department of Housing and Urban Development (HUD): 2011, HUD 4155.1, Mortgage Credit Analysis for Mortgage Insurance: Calculating Maximum Mortgage Amounts on Purchase Transactions.

Zillow: 2018, Consumer Housing Trends Report, *Technical report*.

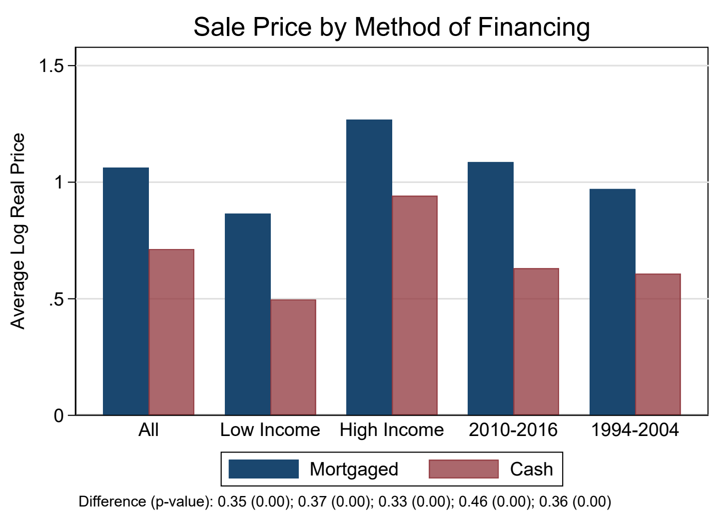
Zillow: 2020, Zillow's Assessor and Real Estate Database (ZTRAX), *Technical report*.

Figures and Tables

Figure 1: Facts About All-Cash Purchases



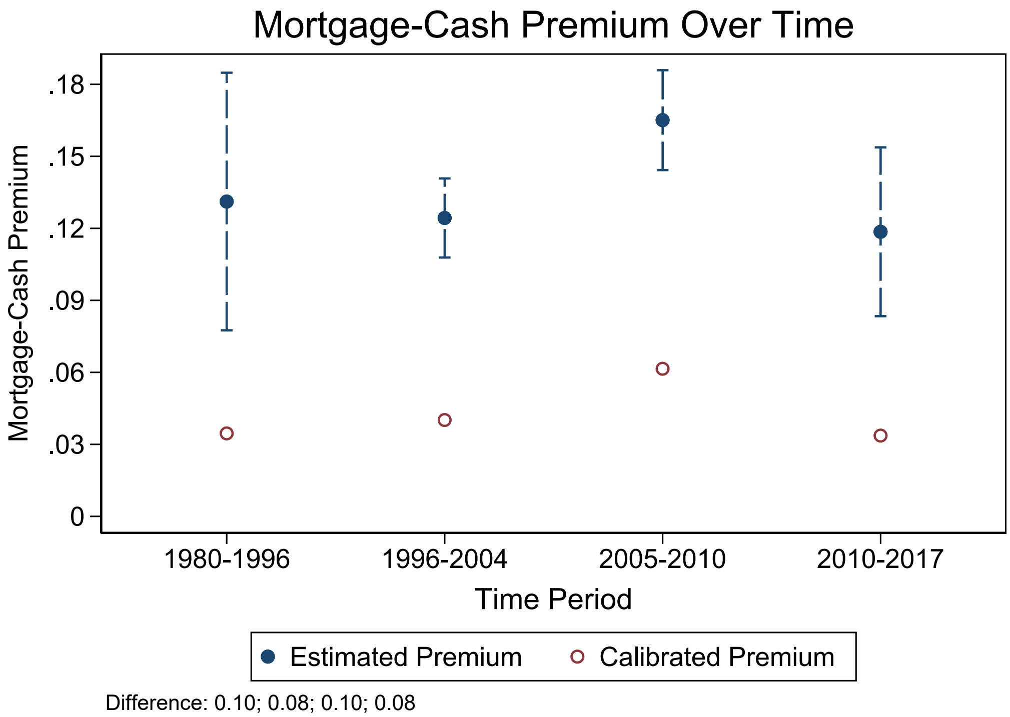
(a) Distribution of Loan-to-Value Ratio



(b) Sale Price by Method of Financing

Note: This figure documents two facts about cash-financed home purchases. Panel (a) plots the distribution of loan-to-value (LTV) ratios across purchases over 1980-2017. Panel (b) plots the average log sales price in hundreds of thousands of 2010 dollars for purchases financed by a mortgage (Mortgaged) and exclusively with cash (Cash). The figure plots this average for different subsamples: All denotes all observed purchases; Low Income and High Income denote purchases in zip codes in the lowest and highest quartile by 2010 income, respectively; and 2010-2016 and 1994-2004 denote purchases over these two periods. Data are from ZTRAX.

Figure 2: Estimated and Calibrated Mortgage-Cash Premium Over Time



Note: This figure plots the empirical and theoretical mortgage cash premium over various time periods. The empirical premium is the value of μ that comes from estimating equation (4) on the subsample $\mathcal{T}$ consisting of months within the indicated time period,

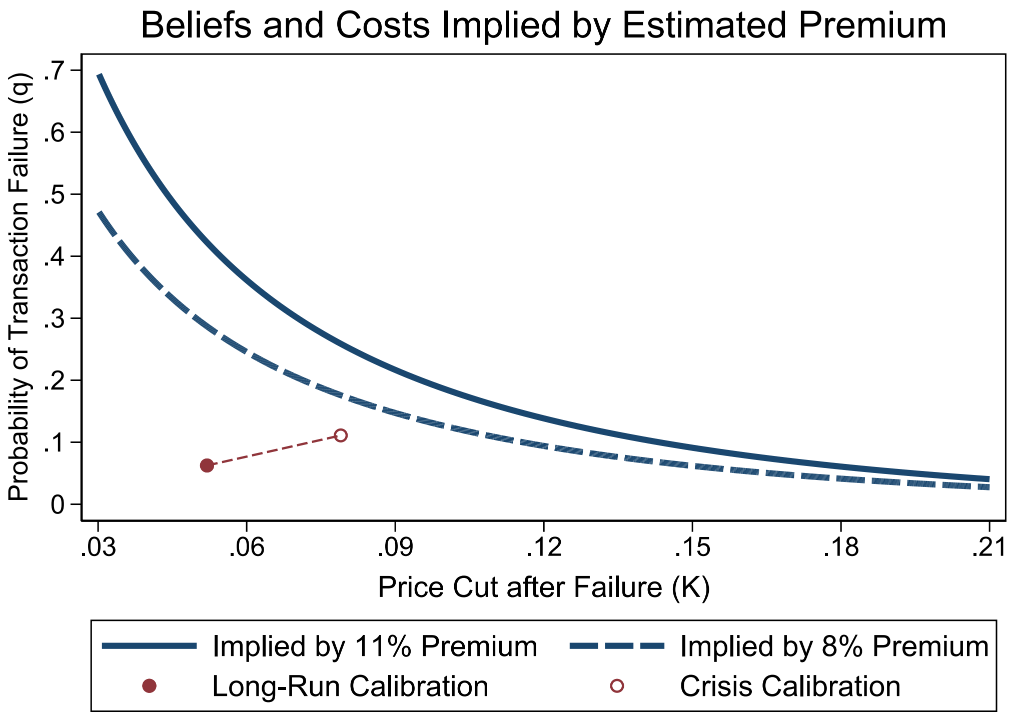
$$\log\left(Price_{i,t}\right) = \mu Mortgaged_{i,t} + \psi X_{i,t} + \zeta_{z(i),t} + \alpha_i + \epsilon_{i,t}, \quad t \in \mathcal{T}$$

where subscripts i and t index property and month; and the remaining terms are defined in the note to Table 2. The theoretical premium is the value of μ that comes from calculating equation (16) for the indicated time period,

$$\mu_{\mathcal{T}} = \xi + \frac{1}{\gamma - 1}\left[q_{\mathcal{T}}\left(\left[\frac{1 - \ell_{\mathcal{T}} - \delta_{\mathcal{T}}}{e^{-\kappa_{\mathcal{T}}} - \ell_{\mathcal{T}} - \delta_{\mathcal{T}}}\right]^{\gamma - 1} - 1\right) - e^{r_{\mathcal{T}}(1-\gamma)} + 1\right],$$

where $q_{\mathcal{T}}$ equals the mortgage application denial rate for years in $\mathcal{T}$, based on data from HMDA; $\kappa_{\mathcal{T}}$ equals the log price discount on forced sales apart from foreclosures for years in $\mathcal{T}$, based on the estimates in Appendix Table A6 of Campbell, Giglio and Pathak (2011); $\ell_{\mathcal{T}}$ and $\delta_{\mathcal{T}}$ are the current loan-to-value ratio and the average simultaneous down payment for years in $\mathcal{T}$, as calculated in Table 7; $r_{\mathcal{T}}$ is the discount rate for years in $\mathcal{T}$, based on the Federal Reserve's discount rate and amortized over a 47 day closing period as in Table 7; and the remaining parameterization is the same as in column (8) of Table 8. The theoretical premium is calculated annually and averaged across years in the time period. For years outside the Campbell, Giglio and Pathak (2011) sample, $\kappa_{\mathcal{T}}$ is parameterized using the log price discount implied by the offer arrival probabilities in Gilbukh and Goldsmith-Pinkham (2019) as described in Appendix C. Brackets are a 95% confidence interval with standard errors clustered by property. The remaining notes are the same as in Tables 2 and 8.

Figure 3: Implied Beliefs and Costs of Transaction Failure

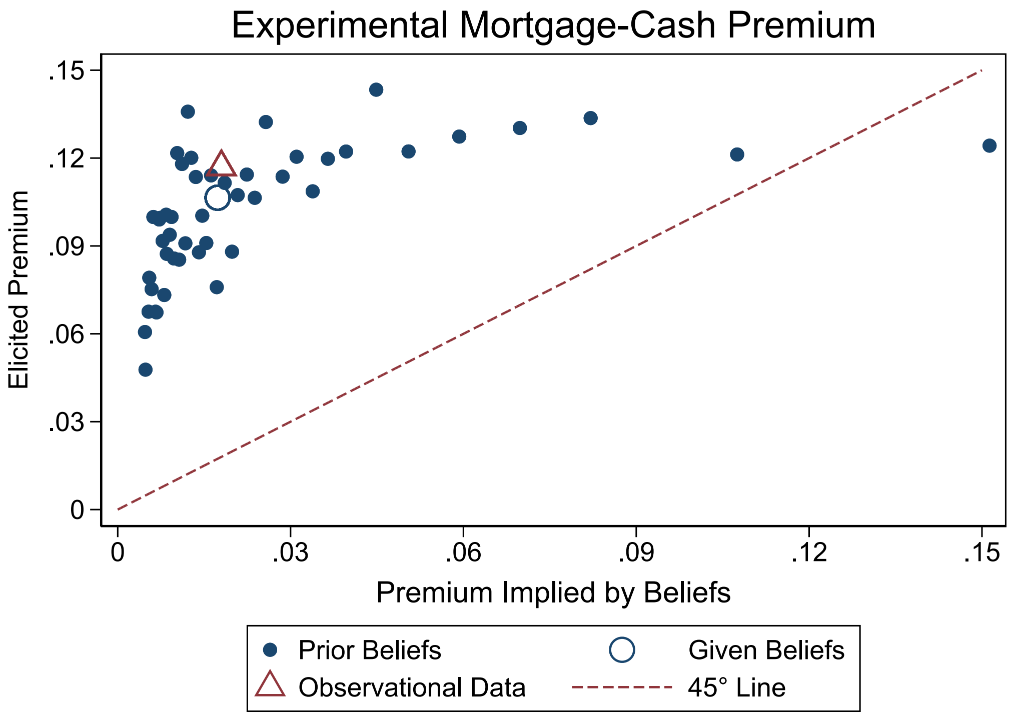


Note: This figure plots the probability of transaction failure and price cut after failure that are consistent with the estimated mortgage-cash premium, according to the extended framework from Section 6.2. Explicitly, the figure plots the relationship

$$q = \left[(\mu - \xi)(\gamma - 1) + e^{r(1-\gamma)} - 1 \right] \left[\left(\frac{1 - \ell - \delta}{e^{-\kappa-\sigma} - \ell - \delta} \right)^{\gamma-1} - 1 \right]^{-1}$$

where q denotes the probability of transaction failure; κ denotes the price cut after failure; μ denotes the mortgage-cash premium; and the remaining parameters are defined in the note to Table 8. The blue solid curve plots the relationship for a mortgage-cash premium of $\mu = 11\%$, and the blue dashed curve plots the relationship for a premium of $\mu = 8\%$. The remaining parameterization is the same as in column (8) of Table 8. The red solid circle plots the long-run values of q and κ from Table 7. The red open circle plots the values from the following crisis parameterization: $q = 11.1\%$, based on the denial rate on pre-approved, first-lien, for-purchase, owner-occupied, single-family mortgage applications in 2008, based on data from HMDA; and $\kappa = 7.9\%$, based on the estimated log price discount on forced sales apart from foreclosures in 2008, based on the estimates in Appendix Table A6 of Campbell, Giglio and Pathak (2011). The remaining notes are the same as in Table 8.

Figure 4: Mortgage-Cash Premium for Survey Respondents

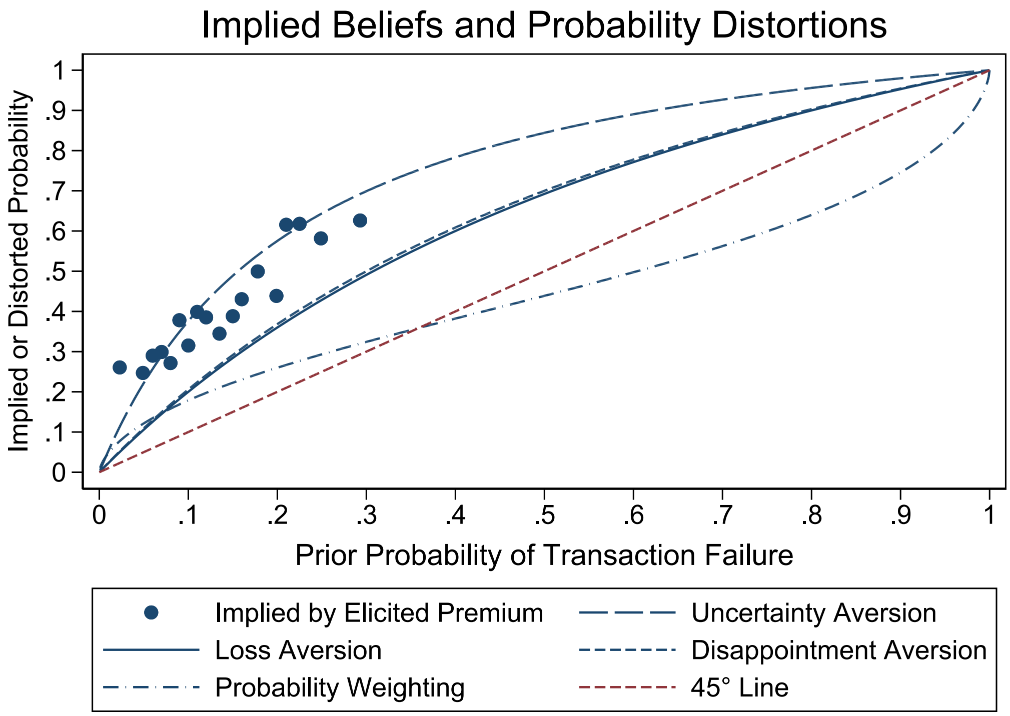


Note: This figure plots survey respondents' observed premium against the theoretical premium implied by their beliefs, based on the extended framework from Section 6. The vertical axis shows the variable $Premium_k$, defined in Table 10 as the percent price premium for a mortgage-financed offer relative to an all-cash offer at which survey respondent k would prefer the mortgage-financed offer, where prices are presented to respondents in both levels and percentages to account for non-proportional thinking. The horizontal axis shows the theoretical mortgage-cash premium, defined in equation (16) as

$$\mu = \xi + \frac{1}{\gamma - 1}\left[q\left(\left[\frac{1-\ell-\delta}{e^{-\kappa-\sigma}-\ell-\delta}\right]^{\gamma-1} - 1\right) - e^{r(1-\gamma)} + 1\right],$$

where q denotes the probability of transaction failure; κ denotes the price cut after failure; μ denotes the mortgage-cash premium; and the remaining parameters are defined in the note to Table 8. The blue solid circles plot this relationship for respondents who are not given values of q and κ, and so it parameterizes these variables according to the percent of the time that the respondent believes a mortgage-financed offer will fail and the price cut that the respondent would impose after a transaction failure, respectively. The blue open circle plots this relationship for respondents who are given values of $q = 7\%$ and $\kappa = 6\%$, and it parameterizes these variables accordingly. The parameters r, γ, σ, and f are parameterized as in Table 7. The remaining mortgage market parameters are parameterized according to the survey's thought experiment: $\xi = 0$; $\ell e^f = 0.30$; and $\delta e^f = 0.15$. For this figure only, the theoretical premium is winsorized at the 5% level so that those respondents with extreme beliefs can fit in the figure. The plot is binned, and each point corresponds to around 20 respondents. The red dashed curve shows the 45-degree line. The coordinates of the red open circle equal the calibrated premium from column (8) of Table 8 along the horizontal axis and the estimated premium from column (1) of Table 2 along the vertical axis. Data are from the survey described in Section 7.

Figure 5: Implied Beliefs and Probability Distortions



Note: This figure plots the probability of transaction failure implied by the elicited mortgage-cash premium among survey respondents and the distorted probability according to the theories described in Section 7. Explicitly, the implied probability equals

$$\hat{q} = \left[(\mu - \xi)(\gamma - 1) + e^{r(1-\gamma)} - 1 \right] \left[\left(\frac{1 - \ell - \delta}{e^{-\kappa - \sigma} - \ell - \delta} \right)^{\gamma - 1} - 1 \right]^{-1},$$

where μ denotes the respondent's elicited premium; and the remaining parameterization is the same as in the note to Figure 4. The blue solid circles plot the implied probability against the respondent's prior probability for respondents who are not given a distribution of outcomes. The blue curves plot the probability used in decision-making as a function of the prior probability, based on the probability distortions described in Section 7. The distortions are uncertainty aversion (e.g., Hansen and Sargent 2011), disappointment aversion (e.g., Gul 1991), loss aversion (e.g., Kőszegi and Rabin 2006), and probability weighting (e.g., Tversky and Kahneman 1992), and the associated expressions for q are

$$q^U = q \left[q + (1 - q) e^{-\theta \left(\frac{u^M - u^R}{(P^C)^{1-\gamma}} \right)} \right]^{-1}, \quad q^D = \frac{q[1 + \beta]}{1 + \beta q}, \quad q^L = \frac{q[1 + \Lambda]}{1 + \Lambda q}, \quad q^W = \frac{q^\alpha}{[q^\alpha + (1 - q)^\alpha]^{\frac{1}{\alpha}}},$$

as shown in equations (C16), (C24), (C28), and (C19), respectively. The figure parameterizes $\theta = 0.33$, $\beta = 1.33$, $\Lambda = 1.25$ and $\alpha = 0.65$ according to columns (3), (5), (7), and (9) of Table 11 and uses the sample average of $\frac{u^M - u^R}{(P^C)^{1-\gamma}}$ to avoid mechanical correlation. The red dashed curve shows the 45-degree line. The remaining notes are the same as in Figure 4.

Table 1: Summary Statistics

Variable	Mean	Standard Deviation	Variable	Mean	Standard Deviation
(a) Transaction-Level Dataset:					
Real Price$_{i,t}$	\$416,813	\$776,315	*Mortgaged*$_{i,t}$	0.642	0.479
Age$_i$	28.431	6.682	*Rooms*$_i$	1.409	1.492
Bathrooms$_i$	0.237	0.715	*Stories*$_i$	1.096	0.111
Air Conditioning$_i$	0.217	0.239	*Detached*$_i$	0.405	0.491
Flip$_{i,t}$	.115	.319	*Foreign Buyer*$_{i,t}$	.003	.054
Same-County Buyer$_{i,t}$	.009	.093	*Institutional Buyer*$_{i,t}$	.001	.038
Cash Propensity$_{b(i,t)}$	.252	.434	*High Seller LTV*$_{i,t}$	0.068	0.252
Same-Month Purchase$_{i,t}$	0.015	0.122	*Foreign Seller*$_{i,t}$	.002	.041
Cash Share$_{s(i,t)}$	0.179	0.327	*Number of Sales*$_{s(i,t)}$	4.687	1.034
(b) Offer-Level Dataset:					
Real Price$_{i,j,t}$	\$512,662	\$359,480	*Mortgaged*$_{i,j,t}$	0.609	0.488
Winning$_{i,j,t}$	0.166	0.372	*Number of Offers*$_{i,j,t}$	5.787	5.120
Baseline Number of Transactions: 426,256					
Baseline Number of Offers: 22,516					

Note: This table summarizes variables from our observational datasets. Panel (a) summarizes variables from our core dataset, the ZTRAX dataset. Subscripts i and t index property and month. Each observation is a home purchase transaction over 1980-2017. The variables are defined as follows: *Real Price*$_{i,t}$ is the sales price in 2010 dollars; *Mortgaged*$_{i,t}$ indicates if the loan amount is positive; *Age*$_i$ is the number of years from when the property was built; *Rooms*$_i$ through *Stories*$_i$ are the number of overall rooms, bathrooms, and stories, respectively; *Air Conditioning*$_i$ and *Detached*$_i$ indicate if i has air conditioning and is a detached single-family home, respectively; *Flip*$_{i,t}$ indicates if the property is subsequently sold within 12 months; *Foreign Buyer*$_{i,t}$ indicates if the buyer has a foreign address; *Same-County Buyer*$_{i,t}$ indicates if the buyer's address is in the same county as the property; *Institutional Buyer*$_{i,t}$ indicates if the buyer is an institution and the property is not to be owner-occupied; *Cash Propensity*$_{b(i,t)}$ indicates whether the buyer of property i in month t, denoted $b(i,t)$, buys another home all-cash over our sample period; *High Seller LTV*$_{i,t}$ indicates if the seller's loan-to-value ratio is above 50%, where the numerator is imputed using a straight-line amortization according to loan term and the denominator is imputed using the median sales price in the buyer's zip code; *Same-Month Purchase*$_{i,t}$ indicates if the seller purchases another home in the same month; *Foreign Seller*$_{i,t}$ indicates if the seller has a foreign address; *Cash Share*$_{s(i,t)}$ is the share of homes sold to cash buyers over our sample period by the seller of property i in month t, denoted $s(i,t)$, after excluding the sale in question and assigning a value of zero to sellers who appear only once in the data; and *Number of Sales*$_{s(i,t)}$ equals the number of sales made by the seller as of month t in the baseline ZTRAX dataset. With the exception of *Mortgaged*$_{i,t}$, all indicator variables are assigned a value of zero when the raw variable is unobserved. Panel (b) summarizes variables from the offer-level dataset. Each observation is an offer to purchase a home over the period from January 2020 through June 2021 made through a real estate agent affiliated with Redfin. The variables are defined as follows: *Mortgaged*$_{i,j,t}$ indicates if j is an offer with a positive loan amount; *Winning*$_{i,j,t}$ indicates if j is the winning offer; *Real Price*$_{i,j,t}$ is the price offered by j in 2010 dollars; and *Number of Offers*$_{i,j,t}$ is the total number of offers, including j. The bottom rows show the number of observations from each dataset that can be used in Tables 2 and 4, respectively. Section 2 and Appendix A contain additional details.

Table 2: Estimated Mortgage-Cash Premium

Outcome:	$\log(Price_{i,t})$										
	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$Mortgaged_{i,t}$	0.191	0.117	0.131	0.124	0.165	0.119	0.090	0.088	0.112	0.108	0.099
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sample	Baseline	Baseline	1980-1996	1996-2004	2005-2010	2010-2017	New Homes	No Flips	Non Instit.	Non Foreign	Positive Equity
Zip Code-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hedonic-Month FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Property FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
R-squared	0.704	0.907	0.969	0.842	0.797	0.826	0.647	0.963	0.642	0.904	0.923
Number of Observations	426,256	426,256	27,087	122,683	69,307	20,792	6,651	186,570	333,288	323,956	313,370

Note: P-values are in parentheses. This table estimates equation (4), which calculates the price premium paid by mortgaged buyers relative to all-cash buyers (i.e., the mortgage-cash premium). Subscripts i and t index property and month. The regression equation is of the form

$$\log(Price_{i,t}) = \mu Mortgaged_{i,t} + \psi X_{i,t} + \zeta_{z(i),t} + \alpha_i + \epsilon_{i,t},$$

where observations are home purchases; $Mortgaged_{i,t}$ indicates if the loan amount is positive; $Price_{i,t}$ is the sales price; α_i is a property fixed effect; $\zeta_{z(i),t}$ is a zip code-by-month fixed effect; and $X_{i,t}$ is a vector of indicators for whether i belongs to bins defined by month and the values of the following hedonic characteristics: the number of years from when the property was built; the number of overall rooms, bathrooms, and stories; an indicator for whether i has air conditioning; and an indicator for whether i is a detached single-family home. The sample period in columns (0)-(1) is 1980-2017. Columns (2)-(5) restrict the sample to months within the indicated time periods. Column (6) restricts the sample to properties built within the previous three years (New Homes). Column (7) restricts the sample to properties that were sold at least 12 months later (No Flips). Column (8) restricts the sample to properties in which neither the buyer nor seller is an institution (Non Instit.). Column (9) restricts the sample to properties in which neither the buyer nor seller has an address outside the U.S. (Non Foreign). Column (10) drops purchases in which the seller's loan-to-value ratio is above 100% (Positive Equity), where the numerator is imputed using a straight-line amortization according to loan term and the denominator is imputed using the average sales price in the surrounding zip code and month to avoid mechanical correlation with $\log(Price_{i,t})$. Standard errors are clustered by property. Data are from the ZTRAX dataset.

Table 3: Summary of Estimates of the Mortgage-Cash Premium

Outcome:	$\log\left(Price_{i,t}\right)$	
	Estimated μ	Table
Repeat-Sales-Hedonic, Baseline	0.117	Table 2, Column 1
Repeat-Sales-Hedonic, 1980-1996	0.131	Table 2, Column 2
Repeat-Sales-Hedonic, 1996-2004	0.124	Table 2, Column 3
Repeat-Sales-Hedonic, 2005-2010	0.165	Table 2, Column 4
Repeat-Sales-Hedonic, 2010-2017	0.119	Table 2, Column 5
Repeat-Sales-Hedonic, New Homes	0.090	Table 2, Column 6
Repeat-Sales-Hedonic, No Flips	0.088	Table 2, Column 7
Repeat-Sales-Hedonic, Non Institutional	0.112	Table 2, Column 8
Repeat-Sales-Hedonic, Non Foreign	0.108	Table 2, Column 9
Repeat-Sales-Hedonic, Positive Equity	0.099	Table 2, Column 10
Non-Accepted Offers	0.086	Table 4, Column 2
Buyer and Seller Characteristics	0.129	Table 5, Column 3
Instrumental Variable	0.139	Table 6, Column 4
Homeowner Survey	0.100	Table 10, Column 2
Hedonic with Non-Repeat Sales	0.161	Appendix Table A1, Column 5
CoreLogic Dataset	0.122	Appendix Table A1, Column 8
Listing Characteristics	0.143	Appendix Table A2, Column 2
Matching	0.169	Appendix Table A3, Column 3
Semi-Structural (Bajari et al. 2012)	0.149	Appendix Table A4, Column 2
Weighting by Representativeness	0.103	Appendix Table A5, Column 2
Repeat-Sales-Hedonic, 61%-80% LTV	0.089	Appendix Figure A2
Repeat-Sales-Hedonic, 81%-90% LTV	0.113	Appendix Figure A2
Repeat-Sales-Hedonic, 91%-125% LTV	0.143	Appendix Figure A2

Note: This table summarizes various estimates of the mortgage-cash premium. Details on each methodology are provided in the notes to the indicated table.

Table 4: Robustness to Using Data on Non-Accepted Offers

Outcome:	$\log\left(Price_{i,j,t}\right)$	
	(1)	(2)
$Mortgaged_{i,j,t} \times Winning_{i,j,t}$	0.081	0.086
	(0.008)	(0.004)
Other Variables:		
$Mortgaged_{i,j,t}$	-0.006	-0.006
	(0.304)	(0.279)
$Winning_{i,j,t}$	-0.044	-0.060
	(0.098)	(0.025)
Month FE	Yes	Yes
Zip FE	Yes	Yes
Offers-on-Property FE	No	Yes
R-squared	0.619	0.620
Number of Observations	22,516	22,516

Note: P-values are in parentheses. This table estimates equation (7), which assesses whether the baseline results are robust to using data on non-accepted offers to estimate the mortgage-cash premium. Subscripts i, j, and t index property, offer, and month. The regression equation is of the form

$$\log\left(Price_{i,j,t}\right) = \mu\left(Mortgaged_{i,j,t} \times Winning_{i,j,t}\right) + \psi_0\, Winning_{i,j,t} + \psi_1 Mortgaged_{i,j,t} + \zeta_{z(i)} + \tau_t + \upsilon_{i,j,t},$$

where observations are home purchase offers; $Mortgaged_{i,j,t}$ indicates if j is an offer with a positive loan amount; $Winning_{i,j,t}$ indicates if j is the offer that is accepted $Price_{i,j,t}$ is the price offered by j; and $\zeta_{z(i)}$ and τ_t are zip code and month fixed effects, respectively. Column (2) includes a vector of fixed effects for the number of offers on the property. The sample consists of purchase offers made through Redfin real estate agents over January 2020 through June 2021. Observations are weighted by the number offers on the property. Standard errors are heteroskedasticity robust. Data are from the Redfin dataset. The remaining notes are the same as in Table 2.

Table 5: Robustness to Controlling for Buyer and Seller Characteristics

Outcome:	$\log\left(Price_{i,t}\right)$		
	(1)	(2)	(3)
$Mortgaged_{i,t}$	0.113	0.100	0.129
	(0.000)	(0.000)	(0.000)
Seller Characteristics:			
$High\ Seller\ LTV_{i,t}$	-0.036	-0.049	-0.047
	(0.000)	(0.000)	(0.000)
$Same\text{-}Month\ Purchase_{i,t}$	0.048	0.048	0.056
	(0.000)	(0.000)	(0.000)
$Foreign\ Seller_{i,t}$	-0.115	-0.125	-0.118
	(0.255)	(0.213)	(0.236)
Buyer Characteristics:			
$Flip_{i,t}$		-0.049	-0.062
		(0.000)	(0.000)
$Foreign\ Buyer_{i,t}$		0.027	0.043
		(0.635)	(0.442)
$Same\text{-}County\ Buyer_{i,t}$		-0.047	-0.039
		(0.002)	(0.009)
$Institutional\ Buyer_{i,t}$		0.204	0.188
		(0.096)	(0.117)
$Cash\ Propensity_{b(i,t)}$			0.101
			(0.000)
Zip Code-Month FE	Yes	Yes	Yes
Hedonic-Month FE	Yes	Yes	Yes
Property FE	Yes	Yes	Yes
R-squared	0.907	0.907	0.908
Number of Observations	426,256	426,256	426,256

Note: P-values are in parentheses. This table estimates a variant of equation (4) that controls for characteristics of the seller and buyer, which accounts for the possibility that parties in all-cash transactions hold lower private valuations than parties in mortgaged transactions. Subscripts i and t index property and month. Column (1) controls for seller characteristics: $High\ Seller\ LTV_{i,t}$ indicates if the seller's loan-to-value ratio is above 50%, where the numerator is imputed using a straight-line amortization according to loan term and the denominator is imputed using the median sales price in the buyer's zip code; $Same\text{-}Month\ Purchase_{i,t}$ indicates if the seller purchases another home in the same month; and $Foreign\ Seller_{i,t}$ indicates if the seller has a foreign address. Column (2) controls for buyer characteristics: $Flip_{i,t}$ indicates if the property is subsequently sold within 12 months; $Foreign\ Buyer_{i,t}$ indicates if the buyer has a foreign address; $Same\text{-}County\ Buyer_{i,t}$ indicates if the buyer's address is in the same county as the property; and $Institutional\ Buyer_{i,t}$ indicates if the buyer is an institution and the property is not to be owner-occupied. Column (3) for $Cash\ Propensity_{b(i,t)}$, defined as an indicator for whether the buyer of property i in month t, denoted $b(i,t)$, buys another home all-cash over our sample period. With the exception of $Mortgaged_{i,t}$, all indicator variables are assigned a value of zero when the raw variable is unobserved. The remaining notes are the same as in Table 2.

Table 6: Robustness to Using Seller Cash Preference as an Instrumental Variable

Outcome:	$Mortgaged_{i,t}$	$\log(Price_{i,t})$	$\log(Price_{i,t})$	$\log(Price_{i,t})$
	(1)	(2)	(3)	(4)
$Mortgaged_{i,t}$		0.116	0.156	0.139
		(0.000)	(0.006)	(0.017)
$Cash\ Share_{s(i,t)}$	-0.058	0.009		
	(0.000)	(0.443)		
Estimator	OLS	OLS	2SLS (Cash Share)	2SLS (Cash Share)
Zip Code-Month FE	Yes	Yes	Yes	Yes
Hedonic-Month FE	Yes	Yes	Yes	Yes
Property FE	No	Yes	Yes	Yes
Seller Controls	No	No	No	Yes
First Stage F-Statistic			313.702	300.770
DWH Statistic (p-value)			0.176	0.363
Number of Observations	425,398	425,395	425,395	425,395

Note: P-values are in parentheses. This table estimates equation (4) after instrumenting for the method of financing (i.e., $Mortgaged_{i,t}$) using the seller's propensity to sell to cash buyers in other purchases. Subscripts i and t index property and month. The instrument $Cash\ Share_{s(i,t)}$ is the share of homes sold to cash buyers over our sample period by the seller of property i in month t, denoted $s(i,t)$, after excluding the sale in question. Sellers who appear only once in the data are assigned a value of zero. Column (1) regresses $Mortgaged_{i,t}$ on $Cash\ Share_{s(i,t)}$ as a first stage. Column (2) estimates a similar specification as in column (1) of Table 2 after controlling separately for $Cash\ Share_{s(i,t)}$, which assesses whether this instrument affects $\log(Price_{i,t})$ only through its effect on through $Mortgaged_{i,t}$, that is, the exclusion restriction in equation (9). The regression equation in columns (3)-(4) is of the same form as in column (1) of Table 2, but it is estimated through 2SLS using $Cash\ Share_{s(i,t)}$ as an instrument for $Mortgaged_{i,t}$. Seller controls are the same as in column (1) of Table 5. Standard errors are clustered by seller. DWH refers to the Durbin-Wu-Hausman test of the null hypothesis that $Mortgaged_{i,t}$ satisfies the exclusion restriction associated with the specifications in Table 2. The remaining notes are the same as in Table 2.

Table 7: Parameters for the Calibrated Mortgage-Cash Premium

Parameter	Value	Source Abbreviation
(a) Mortgage Market Parameters:		
Probability of Transaction Failure (q)	0.060	NAR
	0.082	HMDA
Price Cut after Failure (κ)	0.052	CGP
	0.057	HRY
	0.060	GG
Seller Closing Costs (ξ)	0.013	HUD
Current Loan-to-Value Ratio (ℓe^{-f})	0.331	ZTX
Simultaneous Down Payment (δe^{-f})	0.175	ZTX
Home Search Cost (σ)	0.018	GPSA
(b) Preference Parameters:		
Discount Rate (r), Annualized	0.030	Standard
Coefficient of Relative Risk Aversion (γ)	5	Standard

Note: This table summarizes the parameter values used to calibrate the mortgage-cash premium implied by the framework in Section 6. The third column shows the abbreviated name of the source for each parameter value: NAR refers to the share of transactions that were terminated over 2015-2021, from the National Association of Realtors RCI dataset; HMDA refers to the average denial rate on pre-approved, first-lien, for-purchase, owner-occupied, one-to-four family mortgage applications, based the HMDA dataset; CGP refers to the log price discount on forced sales apart from foreclosures, based on the estimates in Table 2 of Campbell, Giglio and Pathak (2011); HRY refers to the percent discount on foreclosed properties, based on the average estimate in Figure 3 of Harding, Rosenblatt and Yao (2012); GG refers to the log price cut that yields the same utility as a sequence of smaller log price cuts after repeatedly failing to attract an offer, based on the offer arrival probability from Gilbukh and Goldsmith-Pinkham (2019) as described in Appendix C; HUD refers to the product of the maximum-allowable seller contribution to a buyer's closing costs (i.e., 6%) for FHA loans and GSE-backed loans with an 80% loan-to-value ratio, from HUD (2011) and Fannie Mae (2019), multiplied by the share of transactions where the seller makes such a contribution (i.e., 21%), from the National Association of Realtors (2018, 2020); ZTX refers the ZTRAX dataset, which we use to calculate both the average current loan-to-value ratio across sellers and the average down payment for sellers who also purchase a home within 12 months of selling their current home, where the latter statistic is multiplied by 0.61 to reflect the fact that 61% of sellers make such a simultaneous down payment (Zillow 2018); and GPSA refers to housing search costs as calibrated in Table 4 of Guren (2018), page 5 of Piazzesi and Schneider (2009), or Table 1 of Andersen et al. (2021), also multiplied by 0.61 to reflect the fact that 61% of sellers buy a home simultaneously. Real estate agent fees, denoted f in the text but not reported in the table, are parameterized at 6%. The annualized discount rate of 3% corresponds to a 0.4% discount rate over the average closing period of 47 days for mortgaged transactions (Ellie Mae 2012).

Table 8: Calibrated Mortgage-Cash Premium

	Parameterization							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Calibrated Mortgage-Cash Premium (μ)	0.007	0.008	0.008	0.009	0.020	0.022	0.025	0.030
Mortgaged Transaction Parameters:								
Probability of Transaction Failure (q)	0.060	0.060	0.060	0.082	0.060	0.060	0.060	0.060
	(NAR)	(NAR)	(NAR)	(HMDA)	(NAR)	(NAR)	(NAR)	(NAR)
Price Cut after Failure (κ)	0.052	0.057	0.060	0.052	0.052	0.052	0.052	0.052
	(CGP)	(HRY)	(GG)	(CGP)	(CGP)	(CGP)	(CGP)	(CGP)
Seller Closing Costs (ξ)					0.013	0.013	0.013	0.013
					(HUD)	(HUD)	(HUD)	(HUD)
Current Loan-to-Value Ratio (ℓe^{-f})						0.331	0.331	0.331
						(ZTX)	(ZTX)	(ZTX)
Simultaneous Down Payment (δe^{-f})							0.175	0.175
							(ZTX)	(ZTX)
Home Search Cost (σ)								0.018
								(GPS)

Note: This table summarizes the calibrated mortgage-cash premium under various parameterizations of the framework in Section 6. The framework relates the mortgage-cash premium, denoted μ, to the following parameters: the probability that a mortgaged transaction fails (q); the log difference between the sales price after a transaction failure and the sales price associated with an all-cash offer (κ); the seller's contribution to a buyer's closing costs under a mortgaged transaction, expressed as a percent of sales price (ξ); the seller's consumption-equivalent discount rate over the time taken to close a mortgaged transaction (r); the ratio of the seller's current mortgage principal to the sales price under an all-cash offer (ℓe^{-f}); the ratio of the seller's costs of moving to a new home within a three month window of the sale to the sales price under an all-cash offer (δe^{-f}); the cost of searching for a new home after a transaction failure (σ); the seller's coefficient of relative risk aversion (γ); and real estate agent fees paid by the seller (f). Explicitly,

$$\mu = \xi + \frac{1}{\gamma - 1}\left[q\left(\left[\frac{1-\ell-\delta}{e^{-\kappa-\sigma}-\ell-\delta} \right]^{\gamma-1} - 1 \right) - e^{r(1-\gamma)} + 1 \right],$$

as in equation (16). The values of preference parameters r and γ are shown in Table 7 and do not vary across columns. The remaining notes are the same as in Table 7.

Table 9: Variation in the Estimated Premium by Transaction Frictions

Outcome:	$\log(Price_{i,t})$			
	(1)	(2)	(3)	(4)
$Mortgaged_{i,t}$	0.099	0.116	0.082	0.090
	(0.000)	(0.000)	(0.000)	(0.000)
$Mortgaged_{i,t} \times$ *Above-Average* $Friction_{i,t}$	0.019	0.063	0.069	0.168
	(0.002)	(0.000)	(0.000)	(0.000)
Friction Measure	Seller LTV	Seller Down Payment	Zip-Month Transactions	Zip-Month Denial Rate
Model Analogue	ℓ	δ	κ	q
Interaction as Control	Yes	Yes	No	No
Zip Code-Month FE	Yes	Yes	Yes	Yes
Hedonic-Month FE	Yes	Yes	Yes	Yes
Property FE	Yes	Yes	Yes	Yes
R-squared	0.907	0.907	0.907	0.907
Number of Observations	426,256	426,256	426,256	426,256

Note: P-values are in parentheses. This table estimates a variant of equation (4), which assesses heterogeneity in the mortgage-cash premium according to the degree of friction in mortgaged transactions. Subscripts i and t index property and month. The regression equation is similar to that in Table 2 after interacting $Mortgaged_{i,t}$ with an indicator for whether the transaction features an above-average degree of friction, denoted *Above-Average* $Friction_{i,t}$. The measures of *Above-Average* $Friction_{i,t}$ are: an indicator for whether the seller's loan-to-value ratio is above the average (Seller LTV); an indicator for whether the seller's subsequent down payment exceeds the average (Seller Down Payment); an indicator for whether the number of transactions in the surrounding zip code and month in the baseline sample is below the average (Zip-Month Transactions); and an indicator for whether the mortgage application denial rate in the surrounding zip code and month is above the average of approximately 8.5%, based on annual data from HMDA over 2004-2016 (Zip-Month Denial Rate). These variables correspond to the parameters ℓ, δ, κ, and q in the framework from Section 6. The average of the seller's LTV and down payment are calculated using the raw ZTRAX dataset and are conditional on being between zero and one. The numerator of the loan-to-value ratio used to construct the seller's LTV is imputed using a straight-line amortization according to loan term, and the denominator is imputed using the average sales price in the surrounding zip code and month to remove mechanical correlation with $Price_{i,t}$. Columns (1)-(2) control separately for *Above-Average* $Friction_{i,t}$, and columns (3)-(4) do not because their definition of *Above-Average* $Friction_{i,t}$ does not vary within a zip code and month. The remaining notes are the same as in Table 2.

Table 10: Experimental Mortgage-Cash Premium

	Pooled	First Wave	Second Wave		
			Given Distribution	Similar Prior Distribution	All Prior Distributions
	(1)	(2)	(3)	(4)	(5)
Premium$_k$ (μ)	0.104	0.102	0.106	0.120	0.103
Failure Probability$_k$ (q)	0.134	0.135	0.070	0.073	0.133
Price Cut$_k$ (κ)	0.045	0.041	0.060	0.057	0.051
Background Information:					
Household Income$_k$	\$98,353	\$101,318	\$93,081	\$79,535	\$98,775
Age$_k$	41.801	41.407	41.672	36.85	42.59
College Educated$_k$	0.694	0.71	0.649	0.691	0.707
Low Risk Aversion$_k$	0.33	0.357	0.324	0.702	0.290
Low Numeracy$_k$	0.117	0.06	0.167	0.000	0.168
Number of Sales$_k$	1.672	1.672			
Number of Respondents in First Wave: 938					
Number of Respondents in Second Wave: 1,105					

Note: This table summarizes the results of an experimental survey that assesses sellers' preference for all-cash offers. The survey asks respondents to imagine they are selling a home to either a cash or mortgage-financed buyer and then asks them questions about how they would behave in this scenario. Subscript k indexes survey respondent. All respondents are U.S. homeowners. Column (1) summarizes all respondents, column (2) summarizes the first wave, and columns (3)-(5) summarize the second wave separately by: whether the respondent is told the probability of transaction failure and the subsequent price cut (Given Distribution); and whether the respondent is not told these parameters and holds a prior within 1 pp of the given values (Similar Prior Distribution) or any prior (All Prior Distributions). The upper panel summarizes the means of the following variables: *Premium*$_k$ is the percent price premium for a mortgage-financed offer relative to an all-cash offer that makes k indifferent between the two, where prices are presented to respondents in both levels and percentages; *Failure Probability*$_k$ is the percent of the time that k believes a mortgaged transaction will fail; and *Price Cut*$_k$ is the log price cut k would impose after a transaction failure. The lower panel summarizes demographic information and the following background variables: *Low Risk Aversion*$_k$ indicates if k is "usually willing to take risks but sometimes reluctant to do so" or "always willing to take risks", as opposed to "usually reluctant to take risks but sometimes willing to do so" or "never willing to take risks"; *Low Numeracy*$_k$ indicates if k responds with more than 1% error to a calculation of probabilities similar to Lipkus, Samsa and Rimer (2001); and *Number of Sales*$_k$ equals the number of times k has sold a home, plus one for the home in question. Respondents are weighted by minutes spent on the survey. Section 7 and Appendix D contain additional details.

Table 11: Subjective Beliefs and Probability Distortions

	Traditional Framework		Probability Distortions							
	Given Distribution	Prior Distribution	Uncertainty Aversion		Disappointment Aversion		Loss Aversion		Probability Weighting	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Average Premium Gap	0.089	0.072	0.007	0.016	0.045	0.039	0.045	0.066	0.057	0.060
Distortion Parameters:										
Parameter Notation			θ	θ	β	β	Λ	Λ	α	α
Literature Value			0.33	0.30	1.33	1.70	1.25	0.19	0.65	0.75
Literature Source			AHS1	AHS2	BGMT	ABL	TK	BJW	TK	BPK
Value to Explain Gap			0.36	0.36	5.2	5.2	5.2	5.2	None	None

Note: This table summarizes the difference between the empirical and theoretical mortgage-cash premium among respondents in an experimental survey of U.S. homeowners, and it assesses the ability of subjective beliefs and probability distortions to explain this difference. The top row reports the Average Premium Gap, defined as the average difference between a respondent's elicited premium, denoted $Premium_k$ in Table 10, and the respondent's theoretical premium, based on variants of the framework from Section 6. The theoretical premium has the form

$$\mu = \frac{1}{\gamma - 1}\left[q\left(\left[\frac{1-\ell-\delta}{e^{-\kappa-\sigma}-\ell-\delta} \right]^{\gamma-1} - 1 \right) - e^{r(1-\gamma)} + 1 \right],$$

where the notation and parameterization are the same as in Figure 4, except that κ is fixed at the value associated with a given distribution (6%). Column (1) summarizes the gap for respondents facing a given distribution of outcomes, and so $q = 7\%$ (Given Distribution). Column (2) summarizes the gap for respondents who are not given a distribution (Prior Distribution). Columns (3)-(10) summarize the gap after distorting the prior probability, q, using the functional form associated with the indicated probability distortion and the indicated distortion parameter value. The distortions are uncertainty aversion, disappointment aversion, loss aversion, and probability weighting, and the associated expressions for q are

$$q^U = q\left[q + (1-q)\, e^{-\theta\left(\frac{u^M - u^R}{(P^C)^{1-\gamma}} \right)} \right]^{-1}, \quad q^D = \frac{q\,[1+\beta]}{1+\beta q}, \quad q^L = \frac{q\,[1+\Lambda]}{1+\Lambda q}, \quad q^W = \frac{q^\alpha}{\left[q^\alpha + (1-q)^\alpha\right]^{\frac{1}{\alpha}}},$$

as shown in equations (18), (19), (20), and (21), respectively. The sources of the distortion parameter values are: two versions of the calibration described by Anderson, Hansen and Sargent (2003) for θ (AHS1, AHS2); Bonomo et al. (2011) and Ang, Bekaert and Liu (2005) for β (BGMT, ABL); Tversky and Kahneman (1992) and Barberis, Jin and Wang (2021) for Λ (TK, BJW); and Tversky and Kahneman (1992) and Booij, Van Praag and Van De Kuilen (2010) for α (TK, BPK). Columns (3)-(4) calculate q^U using the sample average of $\frac{u^M - u^R}{(P^C)^{1-\gamma}}$ to avoid mechanical correlation. The bottom row shows the value of the indicated distortion parameter that implies an average premium gap of zero. The remaining notes are the same as in Table 10. Section 7.4 and Appendix C contain additional details.

Table 12: Sale Experience as a Proxy for Less Uncertainty

Outcome:	$Premium_k$			$\log(Price_{i,t})$	
	(1)	(2)	(3)	(4)	(5)
$Number\ of\ Sales_k$	-0.010	-0.006	-0.007		
	(0.000)	(0.003)	(0.004)		
$Mortgaged_{i,t} \times Number\ of\ Sales_{s(i,t)}$				-0.005	-0.004
				(0.001)	(0.005)
Other Variables:					
$Failure\ Probability_k$		0.379	0.370		
		(0.000)	(0.000)		
$Price\ Cut_k$		0.070	0.043		
		(0.107)	(0.324)		
$Mortgaged_{i,t}$				0.149	0.143
				(0.000)	(0.000)
$Number\ of\ Sales_{s(i,t)}$				-0.020	-0.020
				(0.000)	(0.000)
Dataset	Experimental Survey			ZTRAX	
Respondent Controls	No	No	Yes		
Seller Controls				No	Yes
Zip Code-Month FE				Yes	Yes
Hedonic-Month FE				Yes	Yes
Property FE				Yes	Yes
R-squared	0.024	0.172	0.239	0.907	0.907
Number of Observations	919	919	918	426,256	426,256

Note: P-values are in parentheses. This table assesses the relationship between the mortgage-cash premium and the seller's experience, a proxy for less uncertainty about mortgaged transactions. Subscripts k, i, and t index survey respondent, property, and month. The datasets used in columns (1)-(3) and (4)-(5) are the first wave of the experimental survey and the baseline ZTRAX dataset, respectively. Columns (1)-(3) estimate a regression equation of the form

$$Premium_k = \psi_0 Number\ of\ Sales_k + \psi_1 X_k + u_k,$$

where observations are U.S. homeowners who partake in the survey; $Premium_k$ is the percent price premium for a mortgage-financed offer relative to an all-cash offer that makes k indifferent between the two; $Number\ of\ Sales_k$ equals the number of times k has sold a home, plus one for the home in question; and X_k is a vector of control variables. The control variables in column (2) are $Failure\ Probability_k$ and $Price\ Cut_k$, defined in the note to Table 10. Column (3) expands the control variables to include a vector of fixed effects for the respondent's state of residence and the variables Age_k through $Low\ Numeracy_k$ shown in Table 10. Standard errors in columns (1)-(3) are heteroskedasticity robust. Columns (4)-(5) estimate a variant of the repeat-sales-hedonic regression equation (4) that interacts $Mortgaged_{i,t}$ with the number of sales made by the seller of property i in month t, denoted $s(i,t)$, as of month t in the baseline ZTRAX dataset. The remaining notes for columns (1)-(3) and columns (4)-(5) are the same as in Table 10 and Table 2, respectively.

Online Appendix

This document contains additional material referenced in the text. Appendix A elaborates on our data description from Section 2. Appendix B performs additional robustness exercises mentioned in the text. Appendix C provides additional details related to the framework from Section 6. Appendix D elaborates on the description of our survey in Section 7.

A Detailed Data Description

We elaborate on the data description in Section 2 by providing additional details about the ZTRAX, offer-level, and additional datasets in Appendices A.1, A.2, and A.3, respectively. We conclude with a catalog of the variables used in our analysis in Appendix A.4.

A.1 ZTRAX Dataset

The ZTRAX dataset is divided into a transactions dataset (ZTrans) and an assessment dataset (ZAsmt). The former dataset contains information on deed transfers, mortgages, and other real estate transactions. The latter dataset contains information on property characteristics from tax assessments. The raw data derive from public records compiled by Zillow and Zillow's proprietary data sources. In Zillow's words: "The Zillow Transaction and Assessment Dataset (ZTRAX) is the nation's largest real estate database made available free of charge to academic, nonprofit and government researchers. ZTRAX is updated quarterly and is continually growing. Released data include: more than 400 million detailed public records across 2,750+ U.S. counties; more than 20 years of deed transfers, mortgages, foreclosures, auctions, property tax delinquencies and more, for both commercial and residential properties; and property characteristics, geographic information and prior valuations for approximately 150 million parcels in 3,100+ counties nationwide" (Zillow 2020). As mentioned in this quotation, Zillow makes ZTRAX available to researchers at no cost, but, in practice, obtaining the data can take between three and twelve months. We obtained the ZTRAX dataset in June 2018 as a single download. For computational convenience, we work with a 25% random sample of the raw dataset that we download. We shall simply refer to this random sample as the "raw ZTRAX dataset".

The raw ZTRAX dataset includes a separate record for each legal document associated with a transaction on a property (e.g. deed transfer, loan document), where properties are identified by parcel number. Therefore, we collapse the raw data to the property-month level. The resulting unit of observation is a home purchase transaction. We then filter the raw data by retaining purchases that satisfy all of the following characteristics:

(a) The property is not in foreclosure, based on the indicator variable used by Zillow to flag such transactions.

(b) The property is not an intra-family transfer, based on the indicator variable used by Zillow to flag such transactions.

(c) The LTV ratio is less than 125%, corresponding to the largest standardized loan product over our sample period (i.e., "125 Loans").

(d) The real sales price exceeds 125% of the gift tax exemption of $35,000, in 2010 dollars.

(e) The real sales price is less than the 99$^{\text{th}}$ percentile across purchase transactions.

(f) The transaction involves only a single property.

(g) There is only a single vesting type.

(h) The property is residential, which includes one-to-four family, condominium, and multi-family properties.

These filters improve the validity of our results because they rule out extreme comparisons that could bias our estimate of the mortgage-cash premium. For example, to the extent that properties in foreclosure or properties owned by a family member are more likely to be cash-financed, we would obtain an upwardly biased estimate. Filters (a), (b), and (d) address these specific cases. The remaining filters further rule out extreme cases or observations highly subject to measurement error. We shall refer to the resulting dataset as the "filtered ZTRAX dataset". The filtered ZTRAX dataset includes 3,528,981 transactions. In terms of coverage, the filtered ZTRAX dataset spans the 1980-2017 period and covers 80% of U.S. counties on a population-weighted basis, or 70% on an unweighted basis. The dataset does not cover the following states: Nevada, Oklahoma, Ohio, North Carolina, New Jersey, or New York. The dataset is most rich over the 1994-2016 period, during which 96% of the purchase transactions in our baseline analysis occur.

We perform our baseline analysis on the subset of transactions in the filtered ZTRAX dataset that can be used to estimate equation (4). Such transactions must satisfy the following conditions: occur in a zip-code-by-month bin in which at least one other transaction occurs; occur on a property that experiences at least one other sale over our sample period; and has information about the hedonic characteristics described below. We refer to the resulting dataset as the "baseline ZTRAX dataset". The baseline ZTRAX dataset includes 426,256 transactions.

We rely on three sets of variables in the ZTRAX dataset. The first and most important set includes the LTV ratio associated with the purchase and its sales price. We describe these two variables in detail in Appendix A.4.

The second set of variables includes information about the buyer and seller in the transaction. We attach an identifier to each seller using the seller's name, and we attach a similar identifier to each buyer. For purchase transactions with multiple sellers or buyers, we define the seller or buyer using the party whose name is listed first. Before creating this identifier, we remove non-alphabetic characters from the name. So, a "seller" is defined as a unique string, and a "buyer" is defined similarly. This methodology will under-classify unique sellers and unique buyers who have common names, or who do not appear first in the deed (e.g., a spouse). It will over-classify unique sellers and unique buyers whose names are spelled differently in the data. We create a separate identifier for whether the seller (buyer) thus-classified is also a buyer (seller) thus-classified within a window around the sale (purchase). We also observe the address of the buyer and indicators for whether the buyer has a foreign address or is institutional. We observe the same information for sellers, although it is less well-populated than for buyers. Lastly, Zillow defines institutions somewhat loosely, as it includes trusts in its definition.

The third set of variables are hedonic characteristics, which come from the ZAsmt component of ZTRAX. We catalog these characteristics in Appendix A.4. The raw data come from tax assessment records, and we assign the characteristics for the most recent assessment to each transaction. We observe hedonic characteristics for 7% of the filtered ZTRAX dataset. To preserve our sample size, we impute unobserved hedonic characteristics using the average of the characteristic within the same zip code-month bin or, when that is not feasible, within the same county-month bin.

A.2 Redfin Dataset

Our offer-level dataset comes from a large U.S. online real estate brokerage, Redfin. We use this dataset in the robustness exercise from Section 5.1. The dataset contains information about offers made on purchase transactions, including both winning and losing offers. The raw data are reported by real estate agents affiliated with Redfin as part of Redfin's Offer Insights program. The Offer Insights program was launched in 2013, but for a number of years the program was limited to certain geographic markets. Therefore, we begin our analysis in 2020, when the Offer Insights program crossed the threshold of covering 50% of U.S. counties. The data used to estimate Table 4 span the period from January 2020 through June 2021.

The unit of observation in the offer-level dataset is an offer for the purchase of a single-family home. We filter the Redfin dataset by retaining offers that satisfy all of the following characteristics:

(a) The associated listing receives at least two total offers.

(b) The zip code of the associated listing has an established Redfin presence, measured by having at least four offers made through a Redfin real estate agent.

(c) The offer comes with a contingency such that the seller cannot recoup the offer price, to match the framework from Section 6.

(d) The offer price lies between 50% and 200% of the list price.

We cannot directly view the microdata out of concern for client privacy. Instead, we conduct our analysis remotely by submitting a program to Redfin's economic analysis division. As mentioned in the text, this workflow discourages data mining.

The information available in the Redfin dataset includes: the month in which the offer was made; the list price of the associated listing; the zip code of the associated listing; a unique property identifier for the associated listing; the size of the property; indicators for whether the property is a detached single-family home, a condominium, or a different type of single-family home (e.g., mobile home); the offer price; an indicator for whether the offer was accepted; the total number of offers made on the property; and the method of financing associated with the offer.

A.3 Additional Datasets

A.3.1 CoreLogic

We use transaction-level data from CoreLogic to assess the external validity of the baseline results in Appendix Table A1. Like Zillow, CoreLogic compiles its data from public records. We use data from CoreLogic's Ownership Transfer dataset. This dataset is analogous to the ZTrans component of the ZTRAX database, except that it only contains information on deed transfers, not mortgages. However, CoreLogic includes an indicator variable in the Ownership Transfer database for whether the transaction is mortgage-financed, which we use to construct $Mortgaged_{i,t}$ as used in Appendix Tabe A1. Consistent with the ZTRAX database, the associated share of mortgage-financed transactions in the CoreLogic dataset equals 65%. Thus, while we do not explicitly observe the loan amount in the CoreLogic database, the financing indicator that we do observe appears to be relatively accurate. We apply the same basic filters described in Appendix A.1 to the CoreLogic database. Since we the Owership Transfer dataset does not include information from tax assessments, the only hedonic characteristic used to construct the hedonic-month fixed effect in Appendix Table A.1 is an indicator for whether the property is a detached single-family home.

A.3.2 National Association of Realtors (NAR)

We use data from the National Association of Realtors (NAR) Realtor Confidence Index Survey (RCI) in the calibration described in Section 6.3. The NAR produces its RCI at the monthly frequency, but it does not make archived issues readily available. The earliest issue of the RCI that we have been able to locate dates back to September 2011. The data are most consistently available over 2015-2021. Over that period, the average annual share of contracts that are terminated equals 6.0%, with an average of 25% of contracts with issues due to financing, 15% with issues due to appraisal, and 14% with issues due to inspection.

A.3.3 Home Mortgage Disclosure Act (HMDA)

We use data on mortgage application denial rates from the Home Mortgage Disclosure Act (HMDA) to parameterize the probability of transaction failure (q) in Section 6.3 and the degree of uncertainty aversion (θ) in Section 7.4. The denial rate is calculated conditional on applications for the purchase (i.e., non-refinance, non-improvement) of an owner-occupied, one-to-four family home that have been pre-approved and that are for a first-lien. To avoid double-counting, we drop all applications for which the action taken is "purchase by institution" (Gete and Reher 2018). To preserve data quality, we drop all applications with a non-empty edit status.

We can only calculate this denial rate over the 2004-2016 period because we only observe pre-approval status and first-lien status over that period. In Figure 2, we do not condition on pre-approval status or first-lien status so that we can parameterize the time-varying failure rate ($q_\mathcal{T}$) consistently over the indicated time periods.

The HMDA data are available at the yearly frequency. In Table 9 and in the calibration of θ described in Appendix C.5, we use zip code-level denial rates. We observe census tracts but not zip codes in HMDA, and so we aggregate the tract-level denial rate to the zip code level using the tract-to-zip code crosswalk from the Department of Housing and Urban Development, weighting tracts by the number of applications. We access the raw HMDA data through Recursion Co.

A.3.4 California Association of Realtors (CAR)

We use data from the California Association of Realtors (CAR) 2019 Seller Consumer Survey in Appendix Table A2. The CAR administered its survey by email to a random sample of consumers throughout California from April 2019 through July 2019. The survey instrument was a questionnaire with both multiple choice and open-ended questions. There were 4,017 valid survey responses, of which 993 had sold a home over the previous 18 months and, thus, are included in our data.

We observe information about the seller's most recent sale. Thus, the unit of observation is the home sale, denoted h in Appendix Table A2. In particular, we observe the following variables: month of sale; county in which the property was sold; sales price; original listing price; days from original listing to sale; number of offers received; whether the property is a single-family detached home; and the age, square feet, and number of bedrooms in the property. We winsorize the sale and listing prices at the 1% level. In estimating Appendix Table A2, we only retain sales with at least two offers and which are on the market for less than 365 days.

We also observe the following background information about the seller: age; gender; race; ethnicity; whether the seller was buying a home at the same time; and whether this was the seller's first time selling a home.

A.3.5 Inflation and Zip Code Income

We use data on average adjusted gross income in 2010 from the IRS SOI Tax Stats to calculate real house price growth by zip code income in Figure 1. Real house prices are in 2010 dollars and are calculated using CPI excluding shelter.

A.4 Catalog of Variables

ZTRAX Dataset

- $Mortgaged_{i,t}$: This variable indicates if the loan amount associated with the purchase of property i in month t is positive.

- $Price_{i,t}$: This variable is the sales price associated with the purchase of property i in month t.

- Age_i: This variable is the number of years from when property i was built.

- $Rooms_i$: This variable is the number of overall rooms in property i.

- $Bathrooms_i$: This variable is the number of overall bathrooms in property i.

- $Stories_i$: This variable is the number of overall stories in property i.

- $Air\ Conditioning_i$: This variable indicates if property i has air conditioning.

- $Detached_i$: This variable indicates if property i is a detached single-family

- $Flip_{i,t}$: This variable indicates if property i is sold within 12 months of its purchase in month t. We assign a value of zero to this indicator variable when the raw variable is unobserved.

- $Foreign\ Buyer_{i,t}$: This variable indicates if the buyer of property i in month t has a foreign address, based on the indicator variable used by Zillow to flag such transactions. We assign a value of zero to this indicator variable when the raw variable is unobserved.

- $Same\text{-}County\ Buyer_{i,t}$: This variable indicates if the buyer of property i in month t has an address in the same county as i. We assign a value of zero to this indicator variable when the raw variable is unobserved.

- $Institutional\ Buyer_{i,t}$: This variable indicates if the buyer of property i in month t is an institution and does not intend to occupy the property, based on the two associated indicator variables used by Zillow to flag such transactions. We assign a value of zero to this indicator variable when the raw variable is unobserved.

- $Cash\ Propensity_{b(i,t)}$: This variable indicates if the buyer of property i in month t, denoted $b(i,t)$, makes another purchase over our sample period in which the home is purchased all-cash.

- $High\ Seller\ LTV_{i,t}$: This variable indicates if the seller of property i has a loan-to-value ratio above 50%. We impute the numerator of the LTV ratio using a straight-line amortization according to loan term. We impute the denominator using the median sales price in the buyer's zip code, which avoids mechanical correlation with $Price_{i,t}$. We assign a value of zero to this indicator variable when the raw variable is unobserved.

- *Same-Month Purchase*$_{i,t}$: This variable indicates if the seller of property i in month t purchases another home in month t. We assign a value of zero to this indicator variable when the raw variable is unobserved.

- *Cash Share*$_{s(i,t)}$: This variable is the share of homes sold to cash buyers over our sample period by the seller of property i in month t, denoted $s(i,t)$, after excluding the sale in question and assigning a value of zero to sellers who appear only once in the data

- *Foreign Seller*$_{i,t}$: This variable indicates if the seller of property i in month t has a foreign address, based on the indicator variable used by Zillow to flag such transactions. We assign a value of zero to this indicator variable when the raw variable is unobserved.

- *Number of Sales*$_{s(i,t)}$: This variable equals the number of sales made by the seller of property i in month t, denoted $s(i,t)$, as of month t in the baseline ZTRAX dataset. We top-code this variable at 5 sales.

Offer-Level Dataset

- *Mortgaged*$_{i,j,t}$: This variable indicates if offer j on property i in month t is an offer with a positive loan amount, based on the offer-level dataset.

- *Winning*$_{i,j,t}$: This variable indicates if offer j on property i in month t is the accepted offer.

- *Price*$_{i,j,t}$: This variable is the price offered by offer j on property i in month t, based on the offer-level dataset. We do not observe the offer price directly, but we observe the list price and the percent above or below price associated with the offer price. We use these two variables to calculate *Price*$_{i,j,t}$.

- *Number of Offers*$_{i,j,t}$: This variable is the number of offers made on property i in month t, including offer j.

B Additional Robustness

We perform additional robustness exercises referenced in Section 5.4.

B.1 Nonparametric Matching Estimator

Two concerns motivate us to estimate the mortgage-cash premium through propensity score matching. First, this nonparametric approach avoids any bias from the linear functional form in equation (4). Second, by construction, a matching estimator ensures that each mortgage-financed purchase has an explicit all-cash counterfactual within the data. This feature supports our coming theoretical assumption that the equilibrium mortgage-cash premium makes sellers indifferent between mortgage and cash-financed offers.

Proceeding in a similar manner as Harding, Rosenblatt and Yao (2012), we construct pairs of home purchases within the same zip code and year according to the probability that the buyer finances the purchase with a mortgage. We calculate this probability, called the "propensity score", through a logistic regression of $Mortgaged_{i,t}$ on the hedonic characteristics from Table 2 and the seller controls from Table 6. Then, we match each mortgaged transaction to a counterfactual all-cash transaction within the same zip code and year.

Panels (a) and (b) of Appendix Table A3 summarize hedonic and seller characteristics of the matched pairs. There are few statistically significant differences between mortgaged and matched all-cash transactions, based on Abadie and Imbens (2006) standard errors, and none of these differences appears economically significant. As noted by Harding, Rosenblatt and Yao (2012), matching within zip codes and years imposes a heavy demand on the data, and so the differences shown in panels (a) and (b) understate the quality of the match.

The estimated mortgage-cash premium equals 16.9%, per the top row of Appendix Table A3. One can interpret this estimate as an "average treatment effect on the treated", as it equals the average difference in log price between mortgage-financed purchases and their counterfactual all-cash match. That we obtain a slightly larger estimate from this nonparametric methodology implies that our baseline estimate does not suffer upward bias from linearity.

B.2 Semi-Structural Estimator

Recall from Section 4.1 that the repeat-sales-hedonic methodology treats the method of financing as if it were a time-varying "hedonic characteristic". This approach leads to an unbiased estimate of the mortgage-cash premium as long as the associated set of fixed effects and controls is sufficiently exhaustive, as stated formally in assumption (8). We relax this assumption by following Bajari et al. (2012) and estimating the mortgage-cash premium semi-structurally.

Bajari et al. (2012) propose a semi-structural methodology for recovering the implicit price of observed characteristics of a transaction (e.g., method of financing) in the presence of time-varying, unobserved characteristics (e.g. a property's corner appeal). The methodology hinges on three assumptions: linearity of the pricing kernel, which we have already assumed in equation (4); Markovian evolution of the unobserved attributes; and rational expectations of market participants.

Building on equation (4), consider a property which transacts in months t and $t+n$. To minimize notation and, more substantively, to avoid the incidental parameters problem, we first residualize the variables $\log(Price_{i,t})$ and $Mortgaged_{i,t}$ against the property and zip code-month fixed effects in our repeat-sales methodology.[20] This step is without loss of generality in terms of obtaining point estimates that are comparable to those in Table 2, per the Frisch-Waugh-Lovell theorem. In particular, the associated pricing kernel is

$$\log(Price_{i,t}) = \mu Mortgaged_{i,t} + \alpha + \epsilon_{i,t}. \tag{B1}$$

Next, suppose the error term $\epsilon_{i,t}$ evolves according to the following Markov process

$$\epsilon_{i,t+n} = \varrho_t(n)\,\epsilon_{i,t} + \omega_{i,t+n}. \tag{B2}$$

Importantly, market participants observe $\epsilon_{i,t}$, but we as econometricians do not. For example, $\epsilon_{i,t}$ may capture the property's curb appeal or the seller's urgency, both of which may correlate with the method of financing but which may affect the transaction price through a separate channel. Equation (B2) introduces some structure in how these attributes evolve over time.

Finally, suppose market participants rely on rational expectations such that

$$\mathbb{E}_t[\omega_{i,t+n}] = 0, \tag{B3}$$

where the expectation is taken with respect to all information available as of month t, as reflected by the month subscript on the expectations operator. In words, equation (B3) states that market participants correctly forecast the value of unobserved attributes of a property's price at the time of its next transaction.

Together, equations (B2) and (B3) provide a moment condition we can use to recover the mortgage-cash premium, μ. Substituting equation (B2) into equation (B1) in month $t+n$ gives

$$\begin{aligned}\log(Price_{i,t+n}) = \mu Mortgaged_{i,t+n} + \alpha + ... \\ ... + \varrho_t(n)\left[\log(Price_{i,t}) - \mu Mortgaged_{i,t} - \alpha\right] + \omega_{i,t+n}.\end{aligned} \tag{B4}$$

Based on the rational expectations assumption in equation (B3), all regressors in equation (B4) are uncorrelated with $\omega_{i,t+n}$ except for the contemporaneous method of financing, $Mortgaged_{i,t+n}$. We address this issue by instrumenting for $Mortgaged_{i,t+n}$ using information

[20]Explicitly, the residualized variables are $\overline{\log(Price_{i,t})} = \log(Price_{i,t}) - \zeta^P_{z(i),t} + \alpha^P_i$ and $\overline{Mortgaged_{i,t}} = Mortgaged_{i,t} - \zeta^M_{z(i),t} + \alpha^M_i$. In words, $\overline{\log(Price_{i,t})}$ and $\overline{Mortgaged_{i,t}}$ are the residuals from a regression of $\log(Price_{i,t})$ and $Mortgaged_{i,t}$ on a vector of zip code-month and property fixed effects. To minimize notation, we continue to denote $\log(Price_{i,t})$ and $Mortgaged_{i,t}$ as such, with the understanding that, within this subsection, they are residualized variables.

available as of month t. Explicitly, the first-stage equation is

$$Mortgaged_{i,t+n} = \bar{\varphi} + \varphi_t^M(n)\, Mortgaged_{i,t} + \varphi_t^P(n) \log(Price_{i,t}) + \nu_{i,t+n}, \tag{B5}$$

where, again making use of rational expectations, $\mathbb{E}_t[\nu_{i,t+n}] = 0$. In practical terms, we estimate equations (B4) and (B5) through a standard two-stage, nonlinear least-squares procedure. We estimate $\varrho_t(n)$, $\varphi_t^P(n)$, and $\varphi_t^M(n)$ as nonparametric functions of the holding period k, rounded to the nearest year.

Appendix Table A4 reports the results. We estimate a mortgage-cash premium of 14.9% using this semi-structural approach, per the result in column (2). Since the sample is necessarily restricted to properties that transact more than twice, we facilitate comparison with the repeat-sales-hedonic results in Table 2 by reestimating equation (4) on this subsample.[21] This results in a similar mortgage-cash premium of 10.9%, shown in column (1).

Along with the instrumental variable results from Section 5.2 and the results based on non-accepted offers from Section 5.1, the semi-structural results from this subsection are quantitatively similar to the estimated mortgage-cash premium of 11% from Table 2. This similarity again supports the validity of the repeat-sales-hedonic methodology.

B.3 Properties without a Repeat Sale

We perform two exercises that assess the external validity of the baseline results with respect to universe of transactions in the ZTRAX dataset.

First, in column (1) of Appendix Table A1, we estimate a premium of 18.6% using the 11,367,195 transactions in the ZTRAX universe that satisfy the basic filters in Appendix A.1. In column (3), we estimate a premium of 12.6%, using the 3,911,805 transactions in the subset of the ZTRAX universe that occur in the zip code-by-month bins that appear in the 25% random sample of the ZTRAX universe that we study in our main analysis, which we call the "filtered ZTRAX dataset" in Appendix A. In column (5), we estimate a premium 16.1%, using the 2,254,389 transactions in the filtered ZTRAX dataset without including a property fixed effect. In column (6), we include a property fixed effect, so that we replicate the specification from column (1) of Table 2. Note that the R-squared equals 58% in column (5), versus 91% in column (6). This finding suggests both that the property fixed effects in equation (4) reduce bias and that the accompanying sample restriction does not jeopardize external validity.

Next, in Appendix Table A5, we reestimate equation (4) after weighting transactions by their inverse probability of appearing in the baseline sample from Table 2, based on observed characteristics (e.g., Solon, Haider and Wooldridge 2015). We obtain the weights through a logistic regression estimated on the filtered ZTRAX dataset, shown in column (1)

[21] In more detail, the number of holding periods that we observe is equal to the number of transactions that we observe minus one. Therefore, we can only include holding period fixed effects (i.e., $\varphi_t^P(n)$, $\varphi_t^M(n)$) for properties that transact at least three times. Otherwise, there is at most only one holding period observed per property, which, given that we have residualized both $\log(Price_{i,t})$ and $Mortgaged_{i,t}$ against property fixed effects, means that there is not enough variation to estimate $\varphi_t^P(n)$ and $\varphi_t^M(n)$.

of Appendix Table A5. The weighted point estimate equals 10.3%, shown in column (2).

B.4 Consistency with the Literature

Two contemporaneous papers have studied the price differential between mortgage-financed and cash-financed home purchases. First, using data on purchases in the Los Angeles MSA between 1999-2017, Han and Hong (2020) estimate a lower premium of 5%. In column (1) of Appendix Table A6, we find a 4.7 pps lower premium when including an interaction term for whether a purchase would fall in the sample studied in Han and Hong (2020). This finding is consistent with their lower estimated premium, which lends external support to our results.

Second, using data on purchases in the Phoenix, Las Vegas, Dallas, and Orlando MSAs and in Gwinnet County, GA between 2013-2018, Buchak et al. (2020) study the price differential between purchases without and with an iBuyer, a specific type of all-cash buyer. This differential equals the difference between the weighted average log price of mortgaged and non-iBuyer all-cash purchases minus the log price of iBuyer purchases. One can recover the mortgage-cash premium implied by this differential by dividing it by the share of non-iBuyer purchases that are mortgage-financed.[22] Since mortgage-financed purchases account for 17% of all purchases in the sample studied by Buchak et al. (2020), as shown in Appendix Table A6, their share of the non-iBuyer market is between 17% and 100%. Accordingly, the 4% iBuyer discount estimated by Buchak et al. (2020) maps to between a 4% and a 21% mortgage-cash premium.

In column (2) of Appendix Table A6, we include an interaction term for whether a purchase would fall in the sample studied in Buchak et al. (2020) and find no evidence that the premium differs for this subsample. Therefore, since our baseline estimate (11%) lies within the range implied by the Buchak et al. (2020) estimate of the iBuyer discount, our two sets of results agree. As with Han and Hong (2020), this similarity supports our results.

[22]Explicitly, let ι denote the iBuyer premium, and let $\mathcal{I}$ and C denote whether a purchase is an iBuyer purchase or a non-iBuyer all-cash purchase, respectively. Suppose that mortgaged purchases account for a share w of all non-iBuyer purchases, and suppose that the mortgage-cash premium is the same for non-iBuyer purchases as for iBuyer ones: $P_{i,t}^{M} = e^{\mu} P_{i,t}^{C} = e^{\mu} P_{i,t}^{\mathcal{I}}$. Then

$$\iota = \mathbb{E}\left[w \log\left(P_{i,t}^{M}\right) + (1-w)\log\left(P_{i,t}^{C}\right)\right] - \mathbb{E}\left[\log\left(P_{i,t}^{\mathcal{I}}\right)\right] = w\mu.$$

Therefore, the implied mortgage-cash premium is $\frac{\iota}{w}$, as stated in the text.

C Mathematical Details

We provide mathematical details related to the framework from Section 6 of the text. In Appendix C.1, we derive the expression for the theoretical mortgage-cash premium. In Appendix C.2, we describe the two micro-foundations for the price cut after transaction failure referenced in Section 6.2 of the text. In Appendix C.3, we derive the comparative statics referenced in Sections 6.1 and 6.2 of the text. Appendix C.4 provides details on the calibration from Section 6.3. In Appendix C.5, we derive expressions for the theoretical mortgage-cash premium according to the non-traditional frameworks described in Section 7.4.

C.1 Derivation of the Theoretical Mortgage-Cash Premium

The mortgage-cash premium implied by the extended framework, μ, comes from solving

$$\frac{\left(P^C_{i,t}e^{-f}-L_{i,t}-D_{i,t}\right)^{1-\gamma}}{1-\gamma}=e^{-\rho}\left[\frac{\left(P^M_{i,t}e^{-a-\xi}-L_{i,t}-D_{i,t}\right)^{1-\gamma}}{(1-q)^{-1}(1-\gamma)}+\frac{\left(P^R_{i,t+1}e^{-f}-L_{i,t}-D_{i,t}\right)^{1-\gamma}}{q^{-1}(1-\gamma)}\right], \tag{C1}$$

which is a variant of equation (13). Rearranging terms in equation (C1) gives

$$(1-\ell-\delta)^{1-\gamma}=e^{-\rho}\left[(1-q)\left(e^{\mu-\xi}-\ell-\delta\right)^{1-\gamma}+q\left(e^{-\kappa-\sigma}-\ell-\delta\right)^{1-\gamma}\right], \tag{C2}$$

which uses the identities $P^M_{i,t}\equiv P^C_{i,t}e^{\mu}$, $P^R_{i,t}\equiv P^C_{i,t}e^{-\kappa}$, $\ell\equiv\frac{L_{i,t}}{P^C_{i,t}e^{-f}}$, and $\delta\equiv\frac{D_{i,t}}{P^C_{i,t}e^{-f}}$. Dividing equation (C2) by $(1-\ell-\delta)^{1-\gamma}e^{-\rho}(1-q)$ gives the following implicit expression for μ,

$$\left(\frac{e^{\mu-\xi}-\ell-\delta}{1-\ell-\delta}\right)^{1-\gamma}=\frac{e^{\rho}}{1-q}-\frac{q}{1-q}\left(\frac{e^{-\kappa-\sigma}-\ell-\delta}{1-\ell-\delta}\right)^{1-\gamma}. \tag{C3}$$

Taking logs of both sides of equation (C3) gives

$$\log\left(\frac{e^{\mu-\xi}-\ell-\delta}{1-\ell-\delta}\right)=\frac{1}{1-\gamma}\log\left[\frac{e^{\rho}-q\left(\frac{e^{-\kappa-\sigma}-\ell-\delta}{1-\ell-\delta}\right)^{1-\gamma}}{1-q}\right]. \tag{C4}$$

Lastly, we apply a first-order approximation to equation (C4) and rearrange terms to obtain

$$\mu\approx\xi+\frac{1}{\gamma-1}\left[q\left(\left(\frac{1-\ell-\delta}{e^{-\kappa-\sigma}-\ell-\delta}\right)^{\gamma-1}-1\right)-e^{r(1-\gamma)}+1\right], \tag{C5}$$

where r is the consumption-equivalent discount rate, defined by solving

$$e^{-\rho}u\left(C\right)=u\left(e^{-r}C\right), \tag{C6}$$

for $u\left(C\right)=\frac{C^{1-\gamma}}{1-\gamma}$. Equation (C5) matches that shown in equation (16) of the text.

C.2 Micro-Foundation for the Price Cut

Column (3) of Table 8 parameterizes κ using a micro-foundation described in Section 6.2. This parameterization complements our baseline parameterizations based on the forced sale discount, shown in columns (1)-(2). We depart from our baseline framework by supposing that sellers experiencing a transaction failure receive an offer in month $t+1$ with probability $\phi < 1$.

Consider an infinitely-lived seller who reduces her log list price by $\tilde{\kappa}$ each month that she fails to receive an offer. As in the text, she immediately accepts the first all-cash offer that arrives. Therefore, her utility in the event of failure equals

$$\begin{aligned}\underline{u}_{i,t}^{M} &= \phi\frac{e^{\tilde{\kappa}(\gamma-1)}}{1-\gamma} + \left(\frac{1-\phi}{e^{\rho}}\right)\left[\phi\frac{e^{2\tilde{\kappa}(\gamma-1)}}{1-\gamma} + \left(\frac{1-\phi}{e^{\rho}}\right)\left[\phi\frac{e^{3\tilde{\kappa}(\gamma-1)}}{1-\gamma}\ldots\right.\right. \\ &= \phi\frac{e^{\tilde{\kappa}(\gamma-1)}}{1-\gamma}\sum_{m=0}^{\infty}\left(\frac{1-\phi}{e^{\rho}}\right)^{m} e^{m\tilde{\kappa}(\gamma-1)} \\ &= \frac{1}{1-\gamma}\frac{\phi e^{\tilde{\kappa}(\gamma-1)}}{1-\left(\frac{1-\phi}{e^{\rho}}\right)e^{\tilde{\kappa}(\gamma-1)}},\end{aligned} \tag{C7}$$

after first dividing by $\left(P_{i,t}^{C}\right)^{1-\gamma}$.

Equation (C7) nests our basic framework, in which $\phi = 1$. Explicitly, our basic framework implies the following expression for a seller's utility in the event of failure

$$\overline{u}_{i,t}^{M} = \frac{e^{\kappa(\gamma-1)}}{1-\gamma} \tag{C8}$$

again after first dividing by $\left(P_{i,t}^{C}\right)^{1-\gamma}$. We solve for the value of κ that equates utility under our basic framework, $\overline{u}_{i,t}^{M}$, with utility under the more-general framework, $\underline{u}_{i,t}^{M}$. Explicitly,

$$\kappa = \frac{1}{\gamma-1}\log\left(\frac{\phi e^{\tilde{\kappa}(\gamma-1)}}{1-\left(\frac{1-\phi}{e^{\rho}}\right)e^{\tilde{\kappa}(\gamma-1)}}\right). \tag{C9}$$

Appendix C.4 describes how we parameterize κ using equation (C9).

C.3 Comparative Statics

We derive the comparative statics related to the basic framework, as discussed in Section 6.1. Then, we derive the comparative statics related to the extended framework, as discussed in Section 6.2.

Basic Framework

Differentiating equation (15) gives the following comparative statics,

$$\frac{\partial \mu}{\partial q} = \frac{e^{\kappa(\gamma-1)} - 1}{\gamma - 1}, \quad \frac{\partial \mu}{\partial \kappa} = q e^{\kappa(\gamma-1)}, \quad \frac{\partial \mu}{\partial r} = e^{r(1-\gamma)}, \tag{C10}$$

all of which are positive. In addition,

$$\begin{aligned}\frac{\partial \mu}{\partial \gamma} &= \frac{(\gamma-1)\left[q\kappa e^{\kappa(\gamma-1)} + r e^{r(1-\gamma)}\right] - \left[q\left(e^{\kappa(\gamma-1)} - 1\right) - e^{r(1-\gamma)} + 1\right]}{(\gamma-1)^2} \\ &= \frac{q e^{\kappa(\gamma-1)}\left[\kappa(\gamma-1) - 1\right] + e^{r(1-\gamma)}\left[1 + (\gamma-1) r\right] - (1-q)}{(\gamma-1)^2}.\end{aligned} \tag{C11}$$

A sufficient condition for equation (C11) to be positive is

$$q e^{\kappa(\gamma-1)}\left[\kappa(\gamma-1) - 1\right] > 1 - q, \tag{C12}$$

substantiating the claim from Section 6.1 that μ is increasing in γ when γ is sufficiently greater than one.

Extended Framework

Differentiating equation (16) gives the following comparative statics,

$$\frac{\partial \mu}{\partial \xi} = 1, \quad \frac{\partial \mu}{\partial \ell} = \frac{\partial \mu}{\partial \delta} = \frac{q\left(1 - e^{-\kappa}\right)}{\left[e^{-\kappa} - \ell - \delta\right]^2}\left(\frac{1 - \ell - \delta}{e^{-\kappa} - \ell - \delta}\right)^{\gamma-2}, \tag{C13}$$

all of which are positive.

C.4 Calibration Details

As stated in Section 6.3, the externally calibrated parameters are q, κ, ξ, ℓ, δ, and σ. These parameters have the following interpretations: the probability of transaction failure (q), the price cut after failure (κ), seller closing costs (ξ), the current loan-to-value ratio (ℓe^{-f}), the simultaneous down payment (δe^{-f}), and the cost of searching for a home (σ). We provide details on the choice of these parameter values as well as how we parameterize the calibrate the mortgage-cash premium over the time periods shown in Figure 2.

Probability of Transaction Failure

The parameterization $q = 6.0\%$ reflects the share of transactions that were terminated over 2015-2021, based on the National Association of Realtors RCI dataset described in Appendix A.3. Repeating from the text, this parameterization may overstate or understate

q. In one direction, q represents the difference in conditional probabilities, and so 6.0% constitutes an understatement because it is an unconditional probability. This understatement is inversely proportional to the share of transactions financed by a mortgage, which, using Table 1, would imply a corrected failure rate of 9.3% and an associated mortgage-cash premium of 0.9% using the model in column (1) of Table 8. In the opposite direction, 6.0% constitutes an overstatement in that cash-financed transactions also fail. However, the most common reasons reported for delayed or terminated settlements are "issues related to obtaining financing" (25%), "appraisal issues" (15%), and "home inspection" (14%), all of which pertain more commonly to mortgage-financed purchases. Thus, correcting for this overstatement would likely result in a failure rate greater than zero but less than 6%.

The alternative parameterization of $q = 8.2\%$ is based on the mortgage application denial rate among first-lien, pre-approved, owner-occupied, one-to-four family home purchase applications over 2004-2016 according to the HMDA data. As described in Appendix A.3, this choice of time period reflects how we only observe pre-approval status and first-lien status over that period. We stated in the text that this parameterization constitutes an upper bound because not all denials as recorded in HMDA necessitate that the buyer terminate the purchase. There are two reasons why a HMDA-recorded denial does not necessitate termination. First, based on discussions with practitioners, originators typically record a mortgage application as "denied" in HMDA if doing so violates their underwriting protocol, but they will subsequently counsel a borrower on how to modify their application so as to be approved in the second round of underwriting. Since the revised application is recorded separately in HMDA, a mortgage-financed transaction that is ultimately successful would nevertheless be associated with a denied application. Second, if borrowers cannot negotiate a second-round application with their initial originator, they can apply for credit from one of its competitors.

Price Cut after Failure

The two parameterizations described in Section 6.3 are $\kappa = 5.2\%$, which comes from Table 2.1 of Campbell, Giglio and Pathak (2011), and $\kappa = 5.7\%$, which comes from Figure 3 of Harding, Rosenblatt and Yao (2012). We also parameterize $\kappa = 6.0\%$ based on the micro-foundation summarized in equation (C9). This value comes from parameterizing $\phi = 55\%$ based on Gilbukh and Goldsmith-Pinkham (2019) and parameterizing $\tilde{\kappa} = 2.9\%$ to match average percent price cut in the U.S. over 2018-2020, the longest history available in Zillow's online database.[23]

Seller Closing Costs

We parameterize $\xi = 0.21 \times 0.06 = 1.3\%$ based on the product of: the share of transactions where the seller makes a contribution to the buyer's closing costs (i.e., 21%) according to NAR (2018, 2020); times the maximum amount a seller can contribute to a buyer's closing costs (i.e., 6%), per the requirements for FHA loans (HUD 2011) and GSE-backed loans

[23]The data can be accessed at: https://www.zillow.com/research/data/.

with a loan-to-value ratio of 80% (Fannie Mae 2019). Since ξ is calculated based on the maximum-allowable seller contribution, it may be interpreted as an upper bound.

Current Loan-to-Value Ratio

We parameterize ℓ directly using the ZTRAX dataset. Explicitly, we parameterize $\ell e^{-f} = 33.1\%$ using the average current loan-to-value ratio across sellers, and then we back out ℓ explicitly using $a = 6\%$.

Simultaneous Down Payment

Similarly, we parameterize δ directly using the ZTRAX dataset. Explicitly, we parameterize $\delta e^{-f} = 17.5\%$ using the average down payment for sellers who also purchase a home within 12 months of selling their current home, which we then multiply by 0.61 to reflect the fact that 61% of sellers make such a simultaneous down payment (Zillow 2018). As with ℓ, we back out δ explicitly using $a = 6\%$.

Home Search Cost

We parameterize σ following the convention in quantitative models of search in the real estate market. For example, Table 4 of Guren (2018) parameterizes a total home search cost of 3%, page 5 of Piazzesi and Schneider (2009) parameterizes a cost of 3.2%, and Table 1 of Andersen et al. (2021) estimates a search cost between 2.4% and 3.5%. We parameterize $\sigma = 1.8$, which comes from multiplying 3% by 0.61 to reflect the fact that 61% of sellers buy a home simultaneously, as just referenced in the context of parameterizing δ.

Time-Varying Parameters

In Figure 2, we calibrate the mortgage-cash premium, μ, over various time periods, $\mathcal{T}$. We can perform this exercise for the parameters q, κ, ℓ, δ, and r. Accordingly: $q_{\mathcal{T}}$ equals the mortgage application denial rate for years in $\mathcal{T}$, based on data from HMDA; $\kappa_{\mathcal{T}}$ equals the log price discount on forced sales apart from foreclosures for years in $\mathcal{T}$, based on the estimates in Appendix Table A6 of Campbell, Giglio and Pathak (2011); $\ell_{\mathcal{T}}$ and $\delta_{\mathcal{T}}$ are the current loan-to-value ratio and the average simultaneous down payment for years in $\mathcal{T}$, which we calculate in the same way as described earlier in this appendix; and $r_{\mathcal{T}}$ is the discount rate for years in $\mathcal{T}$, based on the Federal Reserve's discount rate and amortized over a 47 day closing period as in Table 7. For this exercise, we do not condition the mortgage application denial rate on an application being pre-approved or for a first lien because we do not observe this information in HMDA for years outside the 2004-2016 window. The remaining parameterization is the same as in column (8) of Table 8. We calculate the theoretical premium annually and average it across years in the time period. For years outside the Campbell, Giglio and Pathak (2011) sample, $\kappa_{\mathcal{T}}$ is parameterized using the log price discount implied by the offer arrival probabilities in Gilbukh and Goldsmith-Pinkham (2019) and equation (C9).

C.5 Nontraditional Channels

We derive expressions for the theoretical mortgage-cash premium according to the nontraditional frameworks described in Section 7.4. Each of these frameworks involves a distortion of sellers' subjective probability of failure. Therefore, we adopt the subscript s to index seller, but we no longer maintain the property (i) and month (t) subscripts.

Uncertainty Aversion

We incorporate uncertainty aversion by applying the Hansen and Sargent (2011) framework of decision-making with unknown model parameters ("robustness"). An alternative approach would be to apply the Gilboa and Schmeidler (1989) framework of maxmin expected utility. Hansen et al. (2002) show how the robustness and maxmin treatment of uncertainty aversion can be equivalent after a reinterpretation of parameters.

As in the Caskey (2008) application of uncertainty aversion to the stock market, sellers have heterogeneous beliefs about the probability of failure, but a seller does not know how far her subjective probability lies from the unknown, true probability. Following Hansen and Sargent (2011), she decides between the all-cash and mortgage offer using the following probability,

$$q_s^U = \arg\min_{\tilde{q}} \left\{ (1-\tilde{q})\, u^M + \tilde{q} u^R + \frac{1}{\theta} \mathcal{H}\left(\tilde{q}; q_s\right) \left(P^C\right)^{1-\gamma} \right\}, \tag{C14}$$

where

$$\mathcal{H}\left(\tilde{q}; q_s\right) = \tilde{q} \log\left(\frac{\tilde{q}}{q_s}\right) + (1-\tilde{q}) \log\left(\frac{1-\tilde{q}}{1-q_s}\right) \tag{C15}$$

is the relative entropy of the distribution described by $\tilde{q}$; q_s is the subjective probability held by seller s, typically called her "reference probability"; u^M and u^R are the utility after a successful and unsuccessful mortgage transaction, respectively, as shown in equation (11); and $\left(P^C\right)^{1-\gamma}$ is a normalization ensuring scale independence, as in comparison models (e.g., Maenhout 2004).

The parameter θ governs seller s's aversion to uncertainty: low values of θ imply that s penalizes distributions far from her subjective belief and, thus, has a low aversion to uncertainty; and, conversely, high values of θ correspond to a high aversion to uncertainty. We emphasize that equation (C14) follows directly from the standard treatment of uncertainty aversion in the robustness literature.

Rewriting equation (C14), the probability distortion implied by uncertainty equals

$$q_s^U = q_s \left[q_s + (1 - q_s)\, e^{-\theta\left(\frac{u^M - u^R}{(P^C)^{1-\gamma}}\right)} \right]^{-1}. \tag{C16}$$

The corresponding mortgage-cash premium is defined by

$$\mu_s^U \approx \xi + \frac{1}{\gamma - 1} \left[q_s^U \left(\left(\frac{1 - \ell - \delta}{e^{-\kappa - \sigma} - \ell - \delta} \right)^{\gamma - 1} - 1 \right) - e^{r(1-\gamma)} + 1 \right], \tag{C17}$$

using a similar logic as in Appendix C.1.

Anderson, Hansen and Sargent (2003) describe how to calibrate the degree of uncertainty aversion, θ. Briefly, θ should not exceed the threshold at which a Bayesian seller learning from historical data would reject the null hypothesis that q_s^U exceeds her reference probability, q_s. Let q_s^a denote the critical value at which s rejects the null hypothesis with Type I error a. The corresponding value of θ then equals

$$\theta^a = \frac{\log(1 - q_s) - \log\left(\frac{q_s}{q_s^a} - q_s\right)}{\frac{u^M - u^R}{(P^C)^{1-\gamma}}} \tag{C18}$$

which follows from rearranging terms in equation (C16).

We adopt two measures of q_s^a. First, we suppose seller s with experience in N_s home sales forms a 90% confidence interval around her subjective probability q_s. We parameterize $N_s = 2.67$, which comes from the average number of sales from Table 10 plus one for the home in which the seller now lives. We parameterize $q_s = 13.4\%$ using the average subjective probability in Table 10, after excluding respondents facing a given distribution in column (3). Therefore, we calculate $q_s^a = q_s + z_{0.95}\sqrt{\frac{q_s(1-q_s)}{N_s}} = 0.46$, where $z_{0.95}$ denotes the inverse cumulative standard normal distribution evaluated at 95%. Substituting into equation (C18) gives $\theta^a = 0.33$.

Next, we suppose sellers learn from the distribution of mortgage application denial rates across zip codes. We calculate the mortgage application denial rate at the census tract-by-year level, which we then aggregate to the zip code-by-year level by number of applications. We parameterize q_s^a using the 95$^{\text{th}}$ percentile of denial rates across zip code-years over our core sample period of 1994-2016, weighting by number of applications. Accordingly, we parameterize $q_s^a = 0.41$. Substituting into equation (C18) gives $\theta^a = 0.30$.

Table 11 cites Anderson, Hansen and Sargent (2003) as the source of the parameterizations just described, denoted AHS. We clarify that Anderson, Hansen and Sargent (2003) propose the method for calibrating θ, not the specific parameter values. In fact, parameterizing $\theta \leq 0.50$ would be considered conservative by the literature's standards. For example, Maenhout (2004) describes a parameterization of $\theta = 14$ as "reasonable" in the context of stock market uncertainty, and Barnett, Buchak and Yannelis (2021) parameterize

$\theta = 0.028^{-1} = 36$ in the context of pandemic uncertainty.[24]

Probability Weighting

Probability weighting refers to observation that individuals often overweight small probabilities when making decisions and, symmetrically, underweight large probabilities. This idea has its roots in the psychology literature, and a number of economics papers have proposed specific functional forms that map the probability an individual has in mind onto the probability they use in optimization.

Tversky and Kahneman (1992) propose the following functional form for probability weighting,

$$q_s^W = \frac{q_s^\alpha}{\left[q_s^\alpha + (1-q_s)^\alpha\right]^{\frac{1}{\alpha}}}. \tag{C19}$$

The corresponding mortgage-cash premium equals

$$\mu_s^W \approx \xi + \frac{1}{\gamma - 1}\left[q_s^W\left(\left(\frac{1-\ell-\delta}{e^{-\kappa-\sigma}-\ell-\delta}\right)^{\gamma-1} - 1\right) - e^{r(1-\gamma)} + 1\right]. \tag{C20}$$

Tversky and Kahneman (1992) estimate $\alpha = 0.65$, which also lies close to the midpoint of subsequent estimates summarized by Booij, Van Praag and Van De Kuilen (2010). More recently, Bernheim and Sprenger (2020) estimate $\alpha = 0.71$, and so we view the interval [0.6, 0.75] as a reasonable range. We follow Barberis, Jin and Wang (2021) and parameterize $\alpha = 0.65$ as our baseline.

As mentioned in the text, we also consider the two-parameter function proposed by Prelec (1998),

$$q_s^{W'} = e^{-\bar{\alpha}_0[-\log(q_s)]^{\bar{\alpha}_1}}. \tag{C21}$$

We parameterize $\bar{\alpha}_0 = 1.08$ and $\bar{\alpha}_1 = 0.53$ based on the the most aggressive estimates summarized by Booij, Van Praag and Van De Kuilen (2010), which come from Bleichrodt and Pinto (2000). This parameterization gives an average mortgage-cash premium of $\mu_s^W = 4.7\%$ across survey respondents.

Disappointment Aversion

Gul (1991) models disappointment as occurring when the realized utility of a gamble falls short of the utility from the gamble's certainty equivalent. Disappointment aversion

[24] A mild amount of uncertainty aversion is appropriate in our setting because sellers' subjective probability of failure (13%) does not lie far, in the relative entropy sense, from the least favorable probability that a Bayesian seller would consider (41% to 46%). Therefore, a degree of uncertainty aversion as large as Maenhout (2004) or Barnett, Buchak and Yannelis (2021) would lead to unreasonably large probability distortions that imply a mortgage-cash premium much larger than 11%.

obtains when total utility exhibits a kink at the disappointment threshold. Equation (13) implies that the all-cash offer constitutes the certainty equivalent of the mortgage-financed offer, in equilibrium. Therefore, applying the Gul (1991) model to our setting implies that the mortgage-cash premium solves the following equation,

$$0 = (1 - q_s) \left[u^M - u^C\right] + q_s (1 + \beta) \left[u^R - u_s^C\right], \tag{C22}$$

where $\beta \geq 0$ governs the degree of disappointment aversion. Rearranging terms gives

$$u^C = \left(1 - q_s^D\right) u^M + q_s^D u^R, \tag{C23}$$

where

$$q_s^D = \frac{q_s \left[1 + \beta\right]}{1 + \beta q_s}. \tag{C24}$$

Note that disappointment aversion is equivalent to the particular probability distortion given by equation (C24). We draw this idea from Barseghyan et al. (2013), who make this point in the health insurance market. The corresponding mortgage-cash premium equals

$$\mu_s^D \approx \xi + \frac{1}{\gamma - 1} \left[q_s^D \left(\left(\frac{1 - \ell - \delta}{e^{-\kappa - \sigma} - \ell - \delta}\right)^{\gamma - 1} - 1\right) - e^{r(1-\gamma)} + 1\right]. \tag{C25}$$

We calibrate β following applications of disappointment aversion to the stock market. Specifically, Bonomo et al. (2011) parameterize $\beta = 1.30$, which follows from the estimates in Epstein and Zin (1991). Likewise, Ang, Bekaert and Liu (2005) parameterize $\beta = 0.6^{-1} = 1.70$, as reported in Table 11.

Loss Aversion

Consider the framework of Kőszegi and Rabin (2006) in which individuals have gain-loss utility that enters additively into the traditional utility function. Let $\bar{P}_s \equiv P^C e^{\nu}$ denote the reference price, and $\bar{u}_s$ denote the corresponding isoelastic flow utility.

We consider the case $0 \leq \nu \leq \mu$, so that the seller incurs a loss following a failed mortgage transaction ($-\kappa < \nu$) but not under an all-cash transaction ($0 \leq \nu$). Thus, mortgaged transactions come with the risk of loss relative to the reference point. This ordering will ensure that incorporating loss aversion raises the theoretical mortgage-cash premium. Otherwise, all offers come with either a gain ($\nu \leq -\kappa$) or a loss ($\nu > \mu$), so that loss aversion does not affect the marginal utility associated with a particular method of financing. Or, in the opposite direction, the all-cash offer comes with a certain loss ($\nu > 0$), which would tend to increase the mortgage-cash premium.

Applying the Kőszegi and Rabin (2006) framework to our setting, the mortgage-cash premium solves

$$u^C + \eta \left[u^C - \bar{u}\right] = (1 - q_s) \left[u^M + \eta \left(u^M - \bar{u}\right)\right] + q_s \left[u^R + \eta\lambda \left(u^R - \bar{u}\right)\right], \tag{C26}$$

where $\eta \geq 0$ governs the importance of gain-loss utility; and $\lambda \geq 1$ governs the degree of loss aversion.

We set the home's listing price as the reference price. This assumption accords with how 90% of sellers cite the listing price as their target price, as shown in Table A7. Our survey aligns the all-cash offer price with the listing price, corresponding to $\nu = 0$ and, thus, $u^C = \bar{u}$. Therefore, rearranging terms gives

$$u^C = \left(1 - q_s^L\right) u^M + q_s^L u^R, \tag{C27}$$

where

$$q_s^L = \frac{q_s \left[1 + \Lambda\right]}{1 + \Lambda q_s}, \tag{C28}$$

and $\Lambda \equiv \frac{\eta}{1+\eta} (\lambda - 1)$. Therefore, as with disappointment aversion, loss aversion is equivalent to a particular probability distortion described by equation (C28). The corresponding mortgage-cash premium equals

$$\mu_s^L \approx \xi + \frac{1}{\gamma - 1} \left[q_s^L \left(\left(\frac{1 - \ell - \delta}{e^{-\kappa-\sigma} - \ell - \delta}\right)^{\gamma-1} - 1\right) - e^{r(1-\gamma)} + 1\right]. \tag{C29}$$

We calibrate Λ using the literature on cumulative prospect theory. First, we parameterize $\lambda = 2.25$ and $\frac{\eta}{1+\eta} = 1$, following Tversky and Kahneman (1992). Next, we follow more recent work by Barberis, Jin and Wang (2021), who parameterize $\lambda = 1.5$ and $\frac{\eta}{1+\eta} = 0.38$. Together, these parameterizations imply a range of Λ between 0.20 and 1.25, as reported in Table 11.

D Detailed Survey Description

In this appendix, we elaborate on the description of the survey in Section 7.1. We describe how we administer the survey in Section D.1. In Section D.2, we outline the survey's structure, transcribe the main questions, and explain the choice of wording. The survey itself can be accessed at: https://ucsd.co1.qualtrics.com/jfe/form/SV_0eLteS7os4ULYQ6.

D.1 Survey Administration

We develop the survey using Qualtrics, an online survey design platform. Then, we recruit survey respondents through Prolific, a firm that specializes in helping researchers administer online surveys. Prolific carefully screens its pool of candidate respondents, and it is a competitor to Amazon's MTurk. A number of recent economics papers have recruited survey respondents through MTurk, as summarized by Lian, Ma and Wang (2018), and Casler, Bickel and Hackett (2013) find that the quality of data collected from MTurk respondents resembles that of data collected through laboratory experiments.

For our purposes, Prolific offers two advantages relative to MTurk. First, Prolific allows researchers to recruit participants who satisfy a more specific set of demographic characteristics than does MTurk. This feature enables us to target our survey exclusively to U.S. homeowners. Second, MTurk surveys tend to attract "professional survey respondents" and automated software (e.g., Kennedy et al. 2021). By contrast, Prolific respondents tend to be more representative within a given set of demographic characteristics, partly because Prolific is a newer entrant into the survey recruitment market (e.g., Peer et al. 2017). Consequently, the quality of data collected from Prolific respondents more closely resembles that of data collected through an ideal survey administered by, say, the U.S. Census Bureau.

Survey respondents are paid \$2.00 for completing the survey. and take an average of 6.1 minutes to do so. We administered the survey's first wave in April 2021 and received data from 1,019 respondents. We administered the survey's second wave in November 2021 and received data from 1,202 respondents. All respondents in both surveys are U.S. homeowners. The two waves of the survey comprise two repeated cross-sections, not a panel.

We retain respondents who spend between two-and-a-half and thirty minutes on the survey and are at least twenty years old, which reduces the sample sizes to 938 and 1,105 for the two waves, respectively, as shown in Table 10.

D.2 Survey Transcript

Respondents begin by consenting to anonymously participate in the survey and by providing their Prolific identification number. Then, they are told: *"In what follows, we will describe a hypothetical scenario in which you are selling a home. You will then be asked several questions about how you would respond in this scenario. Please be assured that your responses will be kept completely confidential."* On the next screen, respondents are asked to select the price range in which a typical home in their neighborhood would sell for from among the following ranges: less than \$50,000; between \$50,000 and \$250,000; between \$250,000 and \$500,000; between \$500,000 and \$1,000,000; or greater than \$1,000,000. The respondent's answer determines the level of prices discussed in the subsequent thought experiment. As discussed shortly, this framing technique addresses issues related to non-proportional thinking, and a similar technique is used in the Federal Reserve Bank of New York's Survey of Consumer Expectations (Liu and Palmer 2021). For reference, the share of respondents who fall into these five bins are, respectively: 1.2%; 38.0%; 41.3%; 16.5% and 2.9%. This frequency distribution rather closely matches the analogous distribution of real house prices in the ZTRAX dataset, in which the corresponding shares of transactions are: 1.7%; 43.6%; 32.4%; 15.7%; and 6.6%. This similarity supports the relevance of our survey results for the rest of the paper.

Description of Thought Experiment

The thought experiment takes place over the following six screens. On the first screen, we ask respondents: *"Imagine that you are selling the home in which you now live and are under contract to purchase another home. In other words, you are trying to sell your current home and move to a home you are buying."* The remaining five screens introduce conditions of the home sale that correspond to parameters from the extended framework in Section 6.2.

On the second screen of the thought experiment, we ask respondents: *"Imagine, also, that you bought your current home ten years ago for \$[$0.5 \times B$], and you have now listed it at \$[B]. The remaining mortgage balance that you owe on it is \$[$0.3 \times B$]. So, what you owe on the home is 30% of what you have listed it at."* The value of B depends on the price range in which a typical home in the respondent's neighborhood would sell for, and the values of B for the five bins are, respectively: \$45,000; \$125,000; \$300,000; \$700,000; and \$1,300,000. Thus, respondents are told all quantities in both levels and percentages.[25] For respondents in the second wave, the share of the list price at which the home was bought ten years ago randomly equals either 0.5 or 0.95, with 50% probability for each value. We find no significant difference in results between respondents assigned to these two purchase price factors, which further motivates our choice of the list price as the relevant reference point in Section 7.4. Note that the quantity \$[$0.3 \times B$] corresponds to the variable $L_{i,t}$ in our framework, such that $\ell e^{-f} = 0.3$, to target the value from Table 7.

The third screen introduces the time deadline described in Section 6.2. We tell respondents: *"To make the down payment on the new home, you will need to finalize the sale of your current home, pay off the balance, and then use the money that's left over. The down payment is \$[$0.15 \times B$]. That is 15% of the price at which you have listed your current home. You must make this down payment within [T] weeks."* The parameter T randomly equals either 6 or 8, with 50% probability for each value. To match our theoretical framework, T will also equal the time it will take to close a mortgaged transaction. Accordingly, the values of T are chosen such that the difference between the all-cash and mortgage closing

[25] For example, a household in whose neighborhood homes typically sell for between \$250,000 and \$500,000 would see the following text: *"Imagine, also, that you bought your current home ten years ago for \$150,000, and you have now listed it at \$300,000. The remaining mortgage balance that you owe on it is \$90,000. So, what you owe on the home is 30% of what you have listed it at."*

periods equals one month (Ellie Mae 2012), as we describe shortly. Note that the quantity $[0.15 \times B]$ maps to $D_{i,t}$, such that $\delta e^{-f} = 0.15$ to target the value from Table 7.

The subsequent three screens describe the set of potential homebuyers. On the fourth screen, we tell respondents: *"You receive purchase offers from two potential buyers, neither of which are your family members. Today, you must accept one of these two offers. You will have to decline the other offer, and you cannot keep it as a backup in case the other potential buyer does not follow through."*

The fifth screen describes an all-cash offer: *"The first buyer will pay for the home using their own money, whom we'll call 'Cash Buyer'. If you accept the offer from the Cash Buyer, you can close the transaction any time within 2 weeks. Since the Cash Buyer already has the money to buy your home, there is almost no risk that the Cash Buyer will fail to follow through."* This statement pertains to respondents for whom $T = 6$. When, instead, $T = 8$, we replace the clause "you can close the transaction any time within 2 weeks" with "the transaction will close in 4 weeks". Thus, as just mentioned, the difference between the all-cash and mortgage closing periods equals one month. We introduce this randomization to test for present focus, but, as described in footnote 19, we find no evidence of a difference in premium according to the closing period of the all-cash transaction.

The sixth screen describes a mortgaged offer: *"The second buyer has borrowed money from a mortgage lender, whom we'll call the 'Mortgaged Buyer'. If you accept the offer from the Mortgaged Buyer, it will take [T] weeks to close the transaction. There is a chance that the Mortgaged Buyer will not be able to secure money from their lender by the end of the [T]-week period. If that happens, then you will need to relist your home in [T] weeks."* This statement pertains to respondents in the first wave and in a 50% random sample of the second wave. The remaining half of respondents in the second wave are given a distribution of outcomes for the mortgage transaction: *"The second buyer has borrowed money from a mortgage lender, whom we'll call the 'Mortgaged Buyer'. If you accept the offer from the Mortgaged Buyer, it will take [T] weeks to close the transaction. There is a 7% chance that the Mortgaged Buyer will not be able to secure money from their lender by the end of the [T]-week period. If that happens, then you will need to relist your home in [T] weeks. To attract another offer, you would also need to cut your list price by 6%, that is, reduce your list price to$[0.94 \times B]$ from $[B]."* Note that this distribution features $q = 7\%$, equal to the contemporaneous failure rate based on market data (NAR 2021a, 2021b), and $\kappa = 6\%$, which matches Table 7.

Core Questions

The survey's core consists of several questions that elicit the respondent's mortgage-cash premium (i.e., μ), her motivation for requiring a positive premium, if relevant, and her beliefs about the probability of transaction failure (i.e., q) and the price cut after failure (i.e., κ).

The first core question asks respondents: *"Suppose that both the Mortgaged Buyer and the Cash Buyer offer to pay $[B]. Which offer would you accept today?"* The Cash Buyer's offer dominates the Mortgaged Buyer's offer in our experiment, and so we expect that respondents will answer that they prefer the Cash Buyer. Reassuringly, 97.3% of respondents do so. For a random 50% of participants in the survey's second wave, both buyers offer to pay 95% of the list price. We include this randomization to test for loss aversion relative to the list price. As described in Appendix C.5, loss averse sellers should require a smaller mortgage-cash premium when an all-cash offer comes in under the list price (i.e., reference point), since loss aversion makes such an offer particularly unattractive. However, we find no significant difference in the experimental premium according to the list price. This finding substantiates the structural evidence from Table 11, in which we similarly found little evidence that loss aversion meaningfully contributes to the mortgage-cash premium.

Conditional on preferring the all-cash offer, respondents are then asked: *"Why would you prefer to sell your home to the Cash Buyer rather than the Mortgaged Buyer? Please select the most important reason."*. Respondents can select one option from the following set:

1. *"If the Mortgaged Buyer backs out and I relist at a lower price, I may not have enough money to meet my mortgage and moving expenses."*
2. *"If the Mortgaged Buyer backs out and I relist at a lower price, I will sell the home at a loss relative to my target price."*
3. *"Even if the Mortgaged Buyer would never back out, the Cash Buyer would close more quickly and end the stressful process of selling my home."*
4. *"Other (please describe)"*

Importantly, the order of the first three of these options is randomized, and so the results do not confound tendencies to select options that appear at the top of a list. Conditional on selecting the option related to a target price, respondents are asked: *"What is your target price?*. They can then select one of the following options:

1. *"The price at which I bought my home."*
2. *"The price at which my home is currently listed."*
3. *"Other (please describe)"*

The order of the first two of these options is again randomized. Appendix Table A7 summarizes the responses.

The following screen contains a series of questions that elicit a respondent's mortgage-cash premium in a multiple price list format. Each question asks the respondent whether she would prefer the all-cash versus the mortgaged offer at gradually increasing offer price spreads. The questions are of the form: *"Suppose the Mortgaged Buyer offers to pay \$[$(1+\tilde{\mu})\times B$]. That is [$100\times\tilde{\mu}$]% more than the Cash Buyer. Which offer would you accept now?"* The offer price spreads $\tilde{\mu}$ range from 4% to 28% in increments of 4 pps. We emphasize that respondents are shown the price differentials in both levels and percentages. In the first wave, we proceeded from 5% to 20% in increments of 5 pps, but the similarity of the experimental premiums between the two waves shown in Table 10 suggests that the reduced granularity does not affect the results.

The majority of respondents switch from preferring the all-cash offer to the mortgaged offer once $\tilde{\mu}$ passes some unique threshold. We define the mortgage-cash premium for respondent k as the midpoint between the minimum value of $\tilde{\mu}$ at which the respondent prefers the mortgaged offer and the maximum value of $\tilde{\mu}$ at which the respondent prefers the all-cash offer. Explicitly,

$$\begin{aligned} Premium_k = \frac{1}{2}\left[\min\left\{\tilde{\mu} \middle| \, Mortgaged\ Offer \succ All\text{-}Cash\ Offer\right\} + \ldots\right. \\ \left. \ldots + \max\left\{\tilde{\mu} \middle| \, All\text{-}Cash\ Offer \succ Mortgaged\ Offer\right\}\right]. \end{aligned} \tag{D1}$$

For respondents in the second wave, we assign a mortgage-cash premium of 30% for respondents who still prefer the all-cash offer at a price spread of 28%, such that the mortgage-cash premium either equals zero or ranges from 2% to 30% in increments of 4 pps. Recalling the difference in step size between the two waves, we assign a mortgage-cash premium of 22.5% for respondents in the first wave who still prefer the all-cash offer at a price spread of 20%. Only 7% of respondents require a mortgage-cash premium at the upper bound of their wave's sequence, and so the results are not sensitive to how we define the value of their premium. Lastly, we assign a missing value to the 2% of respondents who exhibit multiple switch points (e.g., Bernheim and Sprenger 2020).

The remaining screens in the survey's core elicit the respondent's beliefs about transaction risk. These screens are only shown to respondents in the first wave or the random 50% of respondents in the second wave who are not given the distribution of mortgage transaction outcomes.

First, we ask respondents: *"What percent of the time do you think the Mortgaged Buyer will back out of the transaction because they fail to secure money from their lender?"* We bound the answers to lie within a range by allowing respondents to select a value between 0% and 30% on a sliding scale. As mentioned in the text, 42% of respondents select a value less than 10% or greater than 20%, suggesting that the scale does not lead to bunching around the midpoint. Let $Failure\ Probability_k$ denote the respondent's prior probability, or, for respondents facing a given distribution, let $Failure\ Probability_k = 7\%$. We then assess numeracy and attentiveness following the commonly used approach of Lipkus, Samsa and Rimer (2001), asking: *"If indeed the Mortgaged Buyer backs out of the transaction [Failure Probability$_k$]% of the time, then how many times out of 1,000 will the Mortgaged Buyer back out?"*. Respondents answer this question by supplying a number.

Similarly, we then ask respondents: *"Homes with lower listing prices typically sell more quickly. At what price would you relist your home if you accept the Mortgaged Buyer's offer today and they subsequently back out?"* Respondents can select a price level between 70% and 100% of the home's list price, again on a sliding scale. Let $\hat{\kappa}_1$ denote the difference between the log of the answer and the log of the list price, or, for respondents facing a given distribution. We again assess numeracy and attentiveness by asking: *"If you indeed need to relist your home at [$e^{-\hat{\kappa}_1}B$] because the transaction fails, by what percent would you need to cut the list price relative to your current list price of [B]?"*. Respondents answer this question by supplying a number, which we denote by $\hat{\kappa}_2$. We define the variable *Price Cut*$_k$ in Table 10 as the average of $\hat{\kappa}_1$ and $\hat{\kappa}_2$, or, when this average exceeds 30%, we define *Price Cut*$_k$ as $\hat{\kappa}_1$. For respondents facing a given distribution, we assess numeracy by simply asking: *"Instead of reducing your list price by 6% if the Mortgaged Buyer backs out, imagine that you reduce your list price by [*10% $\times$ *B]. How large is this [*10% $\times$ *B] reduction relative to your current list price of [B]?".*

We define the variable *Low Numeracy*$_k$ in Table 10, as an indicator that equals one if: (a) the answer to the numeracy calculation posed in terms of the failure rate lies outside 1 pps of the correct answer; or (b1) the answer to the numeracy calculation posed in terms of the price cut lies outside 15 pps of the correct answer and the respondent does not face a given distribution; or (b2) the answer to the numeracy calculation posed in terms of the price cut lies outside 15 pps of the correct answer and the respondent faces a given distribution.

Background Questions

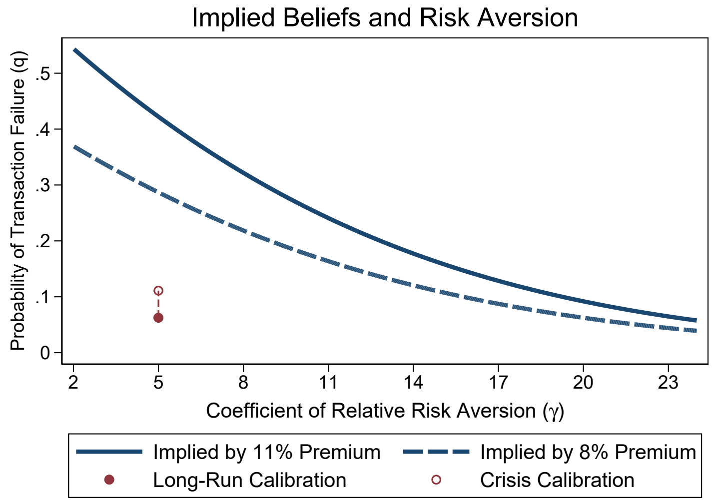
The remainder of the survey asks respondents eleven background questions related to their demographic characteristics, risk attitude, and experience in real estate markets. The first four questions ask the respondent to provide her age, approximate annual household income, state of residence, and highest level of educational attainment. In the fifth question, we follow Fuster and Zafar (2021) and assess a respondent's risk aversion by asking respondents: *"In financial matters, are you generally a person who is willing to take risks, or do you try to avoid taking risks?"* The set of possible answers are:

1. *"I am always willing to take risks."*
2. *"I am usually willing to take risks, but I am sometimes reluctant to do so."*
3. *"I am usually reluctant to take risks, but I am sometimes willing to do so."*
4. *"I am never willing to take risks."*

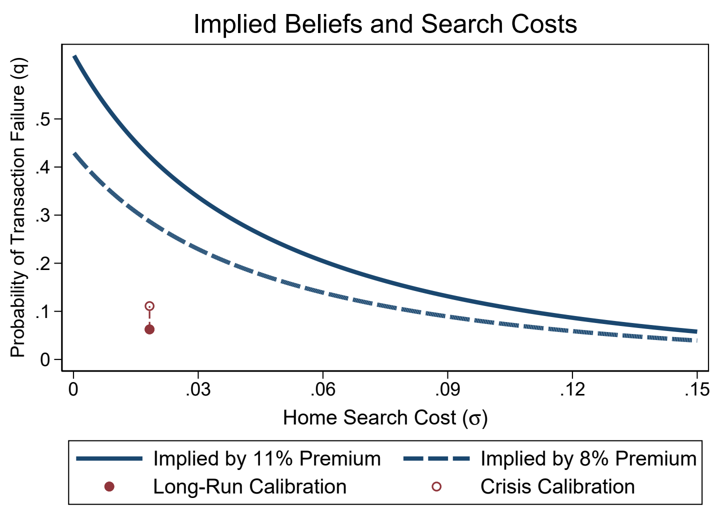
The variable *Low Risk Aversion*$_k$ indicates if respondent k chooses either the first or the second option from this list. In the first wave, we ask respondents whether they have ever sold a home and, if so, the number of occasions.

Additional Figures and Tables

Figure A1: Implied Beliefs by Risk Aversion and Home Search Costs



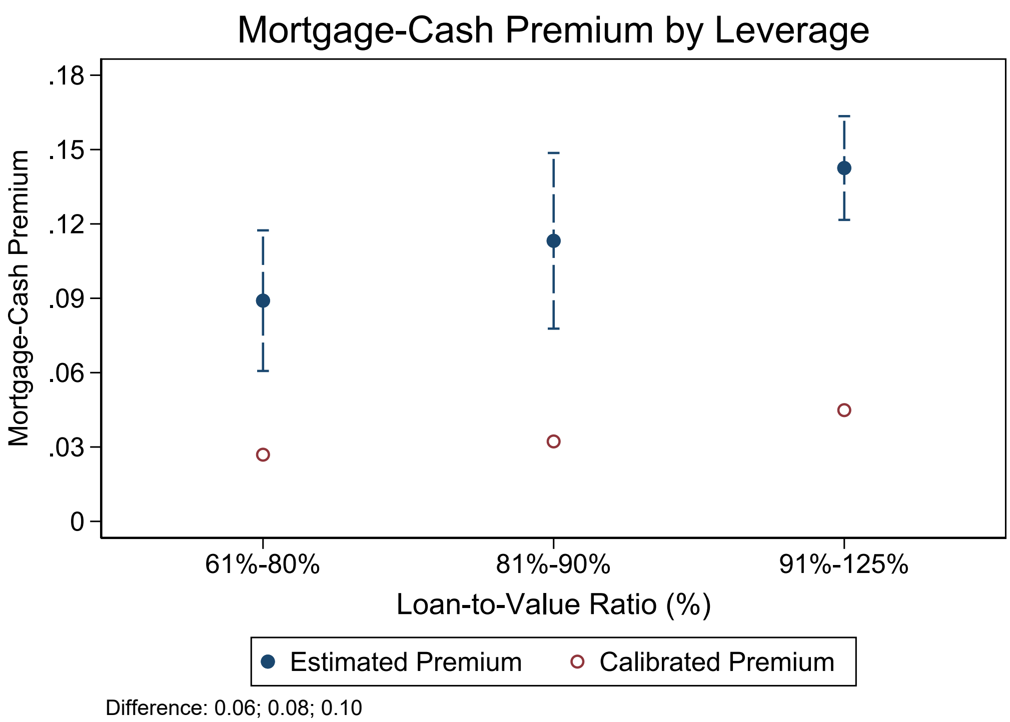
(a) Beliefs Implied by Estimated Discount and Risk Aversion



(b) Beliefs Implied by Estimated Discount and Search Costs

Note: This figure plots the probability of transaction failure consistent with the estimated mortgage-cash premium, according to the extended framework from Section 6.2. The figure is similar to Figure 3, except that it fixes the price cut after failure at its value from column (8) of Table 8. Panel (a) plots the implied probability of failure (q) as a function of the coefficient of relative risk aversion (γ), and panel (b) does so as a function of the home search cost (σ). The remaining notes are the same as in Table 8.

Figure A2: Estimated and Calibrated Mortgage-Cash Premium by Leverage



Note: This figure plots the empirical and theoretical mortgage cash premium over various bins of the loan-to-value (LTV) ratio. The empirical premium is the value of μ that comes from estimating equation (4) on the subsample $\mathcal{L}$ consisting of cash-financed purchases and mortgage-financed purchases within the indicated LTV bin,

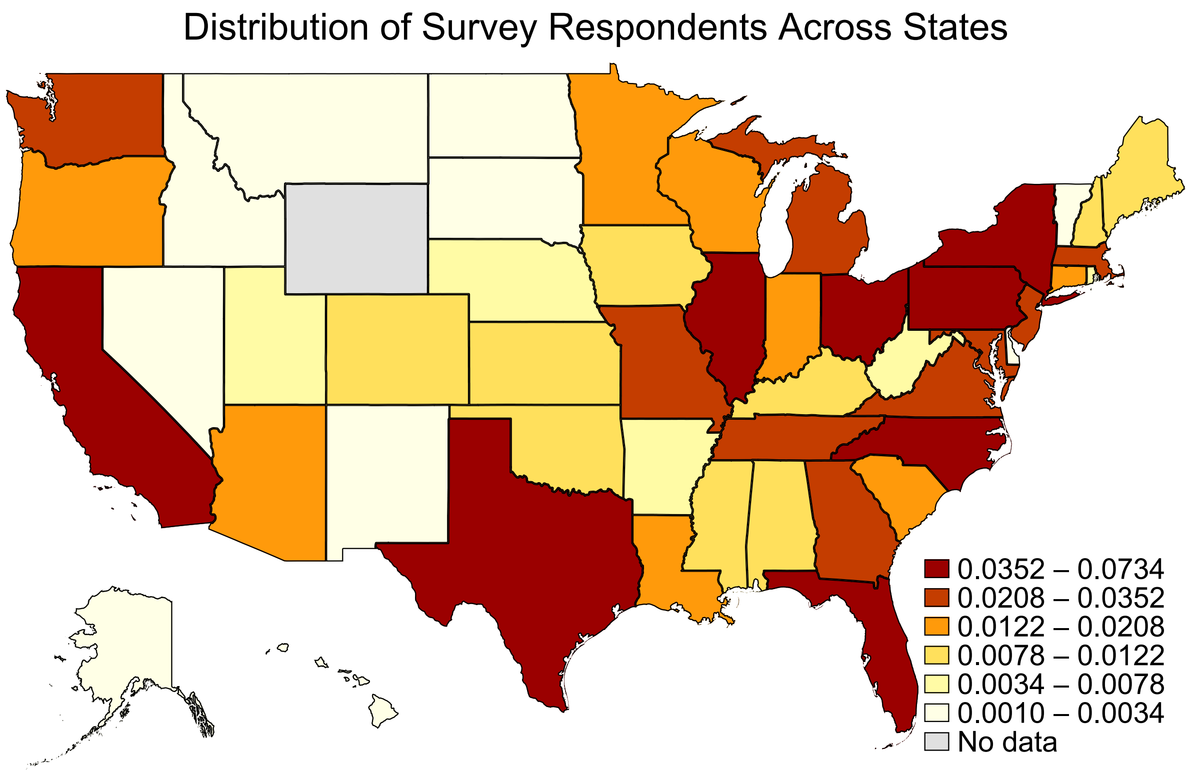
$$\log\left(Price_{i,t}\right) = \mu Mortgaged_{i,t} + \psi X_{i,t} + \zeta_{z(i),t} + \alpha_i + \epsilon_{i,t}, \quad (i,t) \in \mathcal{L}$$

where subscripts i and t index property and month; and the remaining terms are defined in the note to Table 2. To remove the mechanical relationship between LTV ratio and $Price_{i,t}$, the LTV ratio is calculated as the ratio of loan amount to average sales price within the same zip code-month, omitting purchases in which this ratio differs from the actual LTV ratio by more than 25 pps. The theoretical premium is the value of μ that comes from calculating equation (16) for the indicated LTV bin,

$$\mu_{\mathcal{L}} = \xi_{\mathcal{L}} + \frac{1}{\gamma - 1}\left[q_{\mathcal{L}}\left(\left[\frac{1-\ell-\delta}{e^{-\kappa}-\ell-\delta}\right]^{\gamma-1} - 1\right) - e^{r(1-\gamma)} + 1\right],$$

where $q_{\mathcal{L}}$ equals the mortgage denial rate for loan applications in $\mathcal{L}$, based on data from HMDA over 2018-2020, the subperiod in which we observe an application's LTV ratio; $\xi_{\mathcal{L}}$ equals the ratio of total loan origination costs to sale price for loan applications in $\mathcal{L}$, also based on HMDA data over 2018-2020; and the remaining parameterization is the same as in column (8) of Table 8. The remaining notes are the same as Figure 2.

Figure A3: Geographic Distribution of Survey Respondents



Note: This figure plots the share of survey respondents from each U.S. state. Warmer colors correspond to a larger share of survey respondents. Data are from the survey described in Section 7.

Table A1: Robustness to Dataset

Outcome:	$\log(Price_{i,t})$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Mortgaged_{i,t}$	0.186	0.181	0.126	0.126	0.161	0.117	0.209	0.122
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Dataset	ZTRAX Universe		ZTRAX Universe, Core Zip-Months		Core Dataset		CoreLogic	
Zip Code-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hedonic-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Property FE	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.513	0.757	0.566	0.812	0.582	0.907	0.652	0.951
Number of Observations	11,367,195	7,869,239	3,911,805	2,847,146	2,254,389	426,256	103,070,160	62,547,998

Note: P-values are in parentheses. This table estimates equation (4) using various datasets, which assesses the external validity of the baseline results. Subscripts i and t index property and month. All datasets are property-by-month panels that derive from the ZTRAX database. The dataset used in columns (1)-(2) consists of all purchases in the ZTRAX database, after the imposing the basic filters described in Appendix A.1 (ZTRAX Universe). The dataset used in columns (5)-(6) consists of all purchases in a 25% random sample of the ZTRAX Universe, which is the paper's core dataset (Core Dataset). The dataset used in columns (3)-(4) consists of all purchases in the ZTRAX Universe that are also in a zip code-by-month bin that lies in the Core Dataset. The dataset used in columns (7)-(8) consists of all purchases in the CoreLogic database, after the imposing the basic filters described in Appendix A.1 (CoreLogic). The CoreLogic database relies on the same underlying public records as the ZTrans component of the ZTRAX database. The hedonic characteristics used in column (8) is an indicator for whether the property is a detached single-family home, as we describe in Appendix A.3. Columns (2), (4), (6), and (8) include a property fixed effect, while the remaining columns do not. The remaining notes are the same as in Table 2.

Table A2: Robustness to Controlling for Listing Characteristics

Outcome:	$\log(\textit{Sale Price}_{h,t})$			
	(1)	(2)	(3)	(4)
$\textit{Mortgaged}_{h,t}$	0.171	0.143	-0.005	0.068
	(0.062)	(0.029)	(0.963)	(0.419)
$\textit{Mortgaged}_{h,t} \times \textit{First Sale}_{s(h,t)}$			0.381	0.163
			(0.042)	(0.255)
Other Variables:				
$\log(\textit{List Price}_{h,t})$		0.450		0.447
		(0.000)		(0.000)
$\log(\textit{Days on Market}_{h,t})$		0.053		0.054
		(0.027)		(0.025)
Hedonic Controls	Yes	Yes	Yes	Yes
Seller Controls	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
R-squared	0.313	0.631	0.320	0.633
Number of Observations	570	570	570	570

Note: P-values are in parentheses. This table estimates a variant of equation (4) using the California Association of Realtors (CAR) dataset, which assesses the robustness of the baseline results to controlling for characteristics of the listing. Subscripts h and t index home sale and month. The main variables are defined as follows: $\textit{Sale Price}_{h,t}$ is the sale price; $\textit{Mortgaged}_{h,t}$ indicates whether the home was sold to a buyer with an "all cash offer", the "ability to close fastest", or an "offer without contingencies"; $\textit{List Price}_{h,t}$ is the price at which the home was initially listed; $\textit{Days on Market}_{h,t}$ is the number of days from initial listing to sale; and $\textit{First Sale}_{s(h,t)}$ indicates whether the seller, denoted $s(h,t)$, is selling a home for the first time. All specifications include county fixed effects, include month fixed effects, control for $\textit{First Sale}_{s(h,t)}$, and control the following additional seller and hedonic variables: the seller's age; an indicator for whether the seller is female; an indicator for whether the seller is black or Hispanic; an indicator for whether the seller is moving to another home at the same time; an indicator for whether the property is a detached single-family home; and the property's age, log square feet, and number of bedrooms. Columns (3)-(4) are analogous to Table 12 in that they interact the method of financing with a measure of the seller's experience, a proxy for less uncertainty. Standard errors are heteroskedasticity robust. Data are from the CAR dataset. Details on this dataset are in Appendix A.

Table A3: Robustness to Nonparametric Matching Estimator

	Mean of Matched Purchases		
	Mortgaged	Matched Cash	Difference
	(1)	(2)	(3)
$\log(Price_{i,t})$	12.298	12.129	0.169
			(0.000)
Hedonic Characteristics:			
Age_i	28.327	28.320	0.007
			(0.772)
$Rooms_i$	1.308	1.310	0.002
			(0.675)
$Bathrooms_i$	0.178	0.179	0.001
			(0.815)
$Stories_i$	1.096	1.095	0.001
			(0.061)
$Air\ Conditioning_i$	0.193	0.193	0.000
			(0.941)
$Detached_i$	0.185	0.182	0.003
			(0.017)
Seller Characteristics:			
$High\ Seller\ LTV_{i,t}$	0.213	0.208	0.005
			(0.002)
$Same\text{-}Month\ Purchase_{i,t}$	0.004	0.005	0.001
			(0.017)
$Foreign\ Seller_{i,t}$	0.000	0.000	0.000
			(0.285)
Matched on Zip Code	Yes	Yes	Yes
Matched on Year	Yes	Yes	Yes
Number of Observations	140,844	140,844	140,844

Note: P-values are in parentheses. This table estimates the mortgage-cash premium through nonparametric matching, which assesses whether the results are robust to using a nonlinear pricing kernel and to limiting the comparison between highly similar purchases. Subscripts i and t index property and month. Each mortgage-financed purchase is matched to a cash-financed purchase within the same zip code and year based on the hedonic characteristics in Table 2 and seller characteristics in Table 6 using a logistic propensity score. Explicitly, we predict whether a purchase is cash-financed by estimating a logistic regression equation within each zip code-by-year bin, taking the characteristics Age_i through $Foreign\ Seller_{i,t}$ as explanatory variables. Then, we obtain the logistic propensity score as the sum of the predicted probability and the bin's identification code scaled by 100, where the scaling ensures that purchases are matched within the same zip code and year. Finally, each mortgage-financed purchase is matched to one cash-financed purchase, with replacement, using the calculated propensity score. The matching is based on a nearest-neighbor algorithm using the package developed by Leuven and Sianesi (2003). Columns (1) and (2) summarize the mean of the indicated variable across mortgage-financed and matched cash-financed purchases. Column (3) summarizes the mean difference across matches and tests for its statistical significance. Standard errors are as in Abadie and Imbens (2006). The remaining notes are the same as in Table 2.

Table A4: Robustness to Semi-Structural Estimator from Bajari et al. (2012)

Outcome:	$\log\left(Price_{i,t}\right)$	
	(1)	(2)
$Mortgaged_{i,t}$	0.109	0.149
	(0.000)	(0.000)
Estimator	OLS	Semi-Structural
Zip Code-Month FE	Yes	Yes
Hedonic-Month FE	Yes	Yes
Property FE	Yes	Yes
Number of Observations	225,897	225,897

Note: P-values are in parentheses. This table estimates equation (4) using the semi-structural estimator from Bajari et al. (2012). Subscripts i and t index property and month. All of the variables have been residualized against the indicated set of fixed effects, so that these fixed effects are not included in the regression to reduce the number of incidental parameters. Column (1) estimates equation (4) using OLS. Column (2) estimates equation (4) using the Bajari et al. (2012) estimator. Explicitly, column (2) shows the estimated value of μ from the following second-stage regression equation,

$$\log\left(Price_{i,t+n}\right) = \mu \widehat{Mortgaged}_{i,t+n} + \alpha + \varrho_t\left(n\right)\left[\log\left(Price_{i,t}\right) - \mu Mortgaged_{i,t} - \alpha\right] + \omega_{i,t+n},$$

where the first-stage regression equation is

$$\widehat{Mortgaged}_{i,t+n} = \bar{\varphi} + \varphi_t^M\left(n\right) Mortgaged_{i,t} + \varphi_t^P\left(n\right)\log\left(Price_{i,t}\right) + \nu_{i,t+n},$$

and where n denotes holding period; and $\varrho_t\left(n\right), \varphi_t^M\left(n\right)$, and $\varphi_t^P\left(n\right)$ are vectors of holding period-year fixed effects. Holding periods are rounded to the nearest year. The sample includes all properties that transacted at least twice over the sample period, and the smaller sample size relative to Table 2 reflects how $Price_{i,t+n}$ is unobserved for a property's final transaction in the sample. Standard errors are clustered by property in column (1) and bootstrapped in column (2). The remaining notes are the same as in Table 2.

Table A5: Robustness to Weighting by Representativeness

Outcome:	$Baseline_{i,t}$	$\log\left(Price_{i,t}\right)$
	(1)	(2)
$Mortgaged_{i,t}$		0.103
		(0.000)
Characteristics:		
Age_i	-0.027	
$Rooms_i$	0.103	
$Bathrooms_i$	-0.144	
$Stories_i$	0.083	
$Air\ Conditioning_i$	-1.468	
$Detached_i$	0.072	
$High\ Seller\ LTV_{i,t}$	2.117	
$Same\text{-}Month\ Purchase_{i,t}$	0.136	
$Foreign\ Seller_{i,t}$	-0.794	
$Flip_{i,t}$	1.335	
$Foreign\ Buyer_{i,t}$	-0.612	
$Same\text{-}County\ Buyer_{i,t}$	-0.020	
$Institutional\ Buyer_{i,t}$	-1.209	
$Cash\ Propensity_{b(i,t)}$	0.359	
$Sales\text{-}per\text{-}Property_{z(i)}$	-1.209	
$Period\ 1996\text{-}2004_t$	0.362	
$Period\ 2005\text{-}2010_t$	0.320	
$Period\ 2010\text{-}2017_t$	-0.045	
Estimator	Logistic	WLS
Zip Code-Month FE	No	Yes
Hedonic-Month FE	No	Yes
Property FE	No	Yes
R-squared		0.943
Number of Observations	2,709,165	426,256

Note: P-values are in parentheses. This table estimates equation (4) after weighting repeat sales by their representativeness, which assesses the external validity of the baseline results. Subscripts i and t index property and month. Column (1) estimates a logistic regression in which the outcome variable is an indicator for whether the transaction is in the baseline ZTRAX dataset, denoted $Baseline_{i,t}$, and the independent variables are observed characteristics of the transaction: Age_i through $Cash\ Propensity_{b(i,t)}$ are described in the note to Table 1; $Sales\text{-}per\text{-}Property_{z(i)}$ is the total number of home sales in the raw ZTRAX dataset over 1980-2017 in property i's zip code, denoted $z(i)$, divided by the total number of properties in $z(i)$ in the raw ZTRAX dataset; and $Period\ 1996\text{-}2004_t$ through $Period\ 2010\text{-}2017_t$ indicate whether month t lies in the associated time period from Table 2, where 1980-1996 constitutes the reference period. The values shown in column (1) are the coefficient estimates, not the marginal effects, and the p-values associated with each coefficient are not shown to conserve space. Column (2) estimates equation (4) through WLS, where transactions are weighted by the reciprocal probability of appearing in the baseline sample, based on the predictions from column (1). The remaining notes are the same as in Table 2.

Table A6: Consistency with Existing Estimates in the Literature

Outcome:	$\log(Price_{i,t})$	
	(1)	(2)
$Mortgaged_{i,t}$	0.121	0.099
	(0.000)	(0.000)
$Mortgaged_{i,t} \times Subsample_{i,t}$	-0.047	-0.149
	(0.000)	(0.509)
Subsample	HH	BMPS
Subsample Cash Share	0.304	0.829
Zip Code-Month FE	Yes	Yes
R-squared	0.714	0.714
Number of Observations	426,256	426,256

Note: P-values are in parentheses. This table estimates a variant of (4), which assesses the consistency of the baseline results with existing estimates related to the mortgage-cash premium in the literature. Subscripts i and t index property and month. Each column interacts $Mortgaged_{i,t}$ with an indicator for whether the transaction falls within the indicated subsample: HH refers to the subsample of purchases in the Los Angeles MSA between 1999-2017, corresponding to the sample studied in Han and Hong (2020); BMPS refers to the subsample of purchases in the Phoenix, Las Vegas, Dallas, and Orlando MSAs and in Gwinnet County, GA between 2013-2018, corresponding to the sample studied in Buchak et al. (2020). The lower panel summarizes the share of purchases that are cash-financed in each subsample. The mortgage-cash premium estimated by Han and Hong (2020) is 5%. The mortgage-cash premium implied by the estimates in Buchak et al. (2020) is between 4% and 21%, depending on iBuyers' share of cash-financed purchases within the subsample. The remaining notes are the same as in Table 2.

Table A7: Motivation for Preferring All-Cash Offers for Survey Respondents

	Share of Respondents	
	Given Distribution	Subjective Beliefs
	(1)	(2)
"Why would you prefer to sell your home to the Cash Buyer?":		
"If the Mortgaged Buyer backs out and I relist at a lower price, I may not have enough money to meet my mortgage and moving expenses."	0.177	0.136
"Even if the Mortgaged Buyer would never back out, the Cash Buyer would close more quickly and end the stressful process of selling my home."	0.630	0.762
"If the Mortgaged Buyer backs out and I relist at a lower price, I will sell the home at a loss relative to my target price."	0.155	0.050
"Other (please describe)"	0.036	0.050
"What is your target price?":		
"The price at which I bought my home."	0.021	0.022
"The price at which my home is currently listed."	0.904	0.906
"Other (please describe)"	0.075	0.072

Note: This table summarizes the motivation for preferring all-cash offers from an experimental survey of U.S. homeowners. Subscript k indexes survey respondent. The upper panel summarizes responses to a question of why a respondent would prefer an offer from an all-cash buyer relative to a mortgage-financed buyer offering the same price. Respondents are restricted to choosing their most important motivation, and so the shares sum to one across motivations. The lower panel summarizes responses to a question of what the seller's target price is, conditional on selecting the motivation: "... I will sell the home at a loss relative to my target price". Column (1) summarizes respondents who are told the probability of transaction failure and the subsequent price cut (Given Distribution). Column (2) summaries respondents who are not told these parameters (Subjective Beliefs). The sample consists of respondents in the survey's second wave, in which these questions are asked. Respondents are shown each possible answer in a random order, and so the ordering in this table does not correspond to the actual ordering. The remaining notes are the same as in Table 10.