The Mortgage-Cash Premium Puzzle *

Michael Reher† and Rossen Valkanov‡

September 2021

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 the combination of non-standard preferences and standard beliefs best explains the 8% puzzle (11% minus 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, Paul Chiou, Andra Ghent, Sonia Gilbukh, Lu Han, Gregor Matvos, Peter Maxted, Will Mullins, Chris Parsons, Alberto Plazzi, Alessandro Previtero, Alberto Rossi, Yuval Rottenstreich, Rick Townsend, Neng Wang and seminar participants at the SFS Cavalcade, EFA Annual Meeting, AREUEA, PBFEAM, NFA Annual Meeting, and FMA Annual Meeting for comments. We are grateful to Zillow for providing their ZTRAX database under academic terms. We thank Recursion Co for providing their data free-of-charge.

†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, such as 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 accepting a $500,000 all-cash offer, roughly the average in our sample, and a mortgaged offer that is $55,000 larger. We show that the magnitude of this premium far exceeds reasonable compensation for transaction frictions. 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 roughly 11% 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 transaction frictions still only implies a premium of about 3%. The remaining 8% constitutes a puzzle from the standpoint of rational models with frictions. Turning to models in which behavioral channels amplify these frictions, we find that non-standard preferences best explain the puzzlingly 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 address this challenge by using a large and nationally representative dataset on 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 approach. We obtain similar estimates when including homes without a repeat sale, excluding homes that are subsequently flipped, excluding transactions involving an institution, 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). Indeed, based on over ten exercises involving different subsamples, estimators, and datasets, we consistently estimate a mortgage-cash premium of around 11%.

The repeat-sales-hedonic estimator has the advantage of transparency, but, to better understand the variation that identifies the mortgage-cash premium, we also use more complicated estimators. For example, we estimate a 13.9% premium using an instrument for the method of financing: the share of other homes that the seller has sold to all-cash buyers. We construct this instrument at the seller level to avoid bias from property condition. Intuitively, a property in poor condition may attract buyers who typically buy all-cash, but the property's seller was determined in its previous sale when its condition was plausibly different. Consistent with the instrument's validity, it does not affect the sales price after controlling separately for the method of financing. Separately, we estimate premiums of 16.9% using propensity score matching (e.g., Abadie and Imbens 2006) and of 14.9% using the Bajari et al. (2012) semi-structural estimator. The former addresses bias from property condition by restricting the comparison to highly similar transactions within the same zip code and year, and the latter does so by introducing a Markov structure for property condition.

Our results do not depend on only observing equilibrium prices in our core, transaction-level dataset. We also collect 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 explicitly control for the possibility that all-cash buyers seek properties in poor condition, systematically win deals (e.g., sophistication, experience), or hold lower private valuations of real estate. We estimate a premium of 7.9% using this offer-level approach, which, as we describe, holds the same economic significance as our baseline estimate.

The estimated mortgage-cash premium covaries with the degree of mortgage transaction 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 transaction failure and the price cut after 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 or loan type (i.e., FHA-insured). 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 standard ingredients and realistic transaction frictions: risk aversion, time discounting, lengthy mortgage closing periods, mortgage-specific closing costs, a seller's debt overhang, and a seller's down payment on her next home. Reasonable parameter values imply a calibrated mortgage-cash premium of no more than 3%, compared to our estimated premium of 11%. We interpret the inability of a standard, rational model to explain the mortgage-cash premium as a puzzle.

We group candidate explanations of this puzzle into three categories (DellaVigna 2009). First, home sellers may hold non-standard beliefs about the probability of transaction failure and its cost. Second, sellers may hold beliefs consistent with the data, but they do not calculate expected utility as a rational person would. Third, sellers may have non-standard preferences. We assess these explanations by posing our motivating thought experiment to a survey of 1,000 geographically representative U.S. homeowners. Importantly, the survey results reproduce our main finding: the average homeowner would require a 10.0% premium to accept a mortgaged offer over a competing all-cash offer. We infer this premium through a series of pairwise comparisons between offers, presented in both dollars and percents.

Little evidence supports non-standard beliefs as an explanation of the puzzle. For example, based on their stated expectations, homeowners assign a similar probability and loss rate to failure as in the data. While those homeowners with more pessimistic beliefs indeed require a higher mortgage-cash premium, the magnitude of their required premium still exceeds that of a homeowner with standard preferences and the same beliefs by 130% (7 pps). Moreover, that such a puzzling discrepancy arises among the well-educated homeowners in our survey who overwhelmingly pass our test of basic numeracy suggests that non-standard optimization cannot resolve the puzzle either.

By contrast, the survey evidence supports two forms of non-standard preferences: present focus and reference dependence. In favor of present focus, 73% of homeowners cite a longer closing period as the primary drawback of a mortgaged transaction, even if they were "completely certain" the transaction would not fail. Moreover, consistent with a particular form of present focus called present bias (e.g., Laibson 1997; O'Donoghue and Rabin 1999), the mortgage-cash premium does not rise when we randomly increase the experiment's mortgage closing period by two weeks. In favor of reference dependence around non-accepted offers, 44% of respondents prefer all-cash offers to avoid ex-post "regret" after a mortgaged transaction fails. Indeed, using our observational data, we estimate a higher mortgage-cash premium for sellers with a higher value of non-accepted offers.

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. Intuitively, non-standard preferences amplify the effect of transaction frictions that, for a rational home seller, would not seem consequential. 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 this section by situating our contribution within the literature. Section 2 describes our data. Section 3 presents motivating facts. Section 4 discusses identification and estimates the mortgage-cash premium. Section 5 performs robustness. Section 6 calibrates the premium using a rational model. Section 7 discusses experimental evidence and non-standard channels. Section 8 concludes. The online appendix has additional material.

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 the inability of traditional models to explain large credit spreads in bond markets.[1] Unlike these papers, we cannot explain the mortgage-cash premium by appealing to covariance between the costs and probability of "default" (i.e., transaction failure). This leads us to consider behavioral channels. We find little evidence in favor of non-standard belief formation, such as representativeness (e.g., Kahneman and Tversky 1972; Bordalo, Gennaioli and Shleifer 2012, 2013), or non-standard optimization, such as

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

non-proportional thinking (e.g., Shue and Townsend 2021). Instead, we find evidence of non-standard preferences. We specifically complement papers studying the effect of present bias on illiquid savings (e.g., Angeletos et al. 2001; Epper et al. 2020; Laibson, Maxted and Moll 2020), since home sellers' willingness to accept all-cash offers at a steep discount may reduce the value of real estate as a savings vehicle.[2] Our evidence on reference dependence follows a long tradition documenting its role in real estate (e.g., Genesove and Mayer 2001; Andersen et al. 2021).

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. In this vein, we relate most closely to contemporaneous and independent work by Han and Hong (2020) and Buchak et al. (2020), who, respectively, document price discounts for all-cash buyers in Los Angeles and for a specific type of all-cash buyer, iBuyers. We differ from these contemporaneous papers by estimating this price discount over a broader sample of all-cash buyers, by assessing the magnitude of this discount, and by evaluating behavioral explanations for it. After accounting for differences in sample, we show how our estimated mortgage-cash premium agrees with the estimates in these papers.

[2] See Frederick, Loewenstein and O'Donoghue (2002), Ericson and Laibson (2019), or Cohen et al. (2020) for summaries of the broader literature on time preferences.

[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.

2 Data

Our analysis relies on two observational datasets and one experimental dataset, the last of which we describe in Section 7. Table 1 summarizes key variables from the two observational datasets. Appendix A provides further details and a catalog of variables.

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. We collapse the ZTRAX dataset into an unbalanced panel across properties i and months t. 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. These filters ensure that we avoid misleading comparisons that could otherwise bias the estimates. The resulting dataset spans 80% of U.S. counties on a population-weighted basis. We prioritize internal validity and perform our baseline analysis on the subsample of properties for which we can include a property fixed effect. Section 5.4 relaxes this restriction.

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. 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. We use this offer-level dataset to assess the internal validity of the results from our transaction-level ZTRAX dataset, as described in Section 5.3. The raw data are reported by real estate agents affiliated with a large U.S. real estate brokerage and cover 53% of U.S. counties. We observe the LTV ratio and sales price associated with each offer, as well as the list price and the number of competing offers, both of which serve as useful controls. Out of respect for clients' privacy, we do not observe the address of the property. Lastly, the offer-level dataset spans May 2020 through October 2020 because the brokerage only retains recent offers. Thus, while the offer-level dataset enhances our results' internal validity by providing information on non-accepted offers, it limits external validity by covering a relatively small subset of the time series and cross section.

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.[5]

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, again based on the ZTRAX dataset. Beginning with the leftmost column, the average difference in real sales price is 35 log points. Moving right, we 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. Appendix Figure A1 replicates Figure 1 using the offer-level dataset.

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.

[5]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).

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

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

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 $P_{i,t}^M$ for cash-financed purchases (i.e., $\mathcal{M}_{i,t} = 0$), and, similarly, we do not observe $P_{i,t}^C$ 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.[6]

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 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.

[6]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.

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, when expanding the sample to include properties without a repeat sale, we obtain a lower R-squared of 58% and a larger point estimate of 16.1%, as shown in Appendix Table A2. Likewise, we estimate a premium of 19.1% when only including zip code-by-month fixed effects in column (0) of Table 2. These findings suggest that the property fixed effects in equation (4) reduce bias, and the accompanying sample restriction does not jeopardize external validity.

Columns (2)-(8) report the estimated premium within various subsamples. We obtain a consistent premium after partitioning the sample by time period, as shown in columns (2)-(5). The higher premium over 2005-2010 corresponds to elevated transaction frictions during that period, as we later discuss in Section 6.4 in the context of Figure 2.

Next, we restrict the sample to transactions with less asymmetric information about the property's unobserved 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.

Lastly, in columns (8)-(9) we estimate premiums of 11.2% and 10.8% in transactions between non-institutional and non-foreign parties, respectively. These two findings imply that we do not counfound 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).

Collecting the previous two paragraphs, we obtain a similar premium of between 9% and 12% when restricting the sample to purchases with relatively few information asymmetries. Therefore, even if these more-transparent purchases are less likely to be financed with cash, as Appendix Table A1 suggests, such sample selection does not appear to bias the estimate.

5 Validity of Identification

We take the question of internal validity very seriously, and so we perform over ten exercises designed to address particular violations of assumption (5). Broadly, we can group the set of potential violations into two categories: those that stem from unobserved characteristics of the property; and those that stem from unobserved characteristics of the buyer and seller exchanging it.

We have already taken a first pass at assessing bias from unobserved, time-varying characteristics in columns (6)-(9) of Table 2, where we restrict the sample to transactions with little asymmetric information. We further assess such bias by: instrumenting for the method of financing using the seller's preference for all-cash offers as revealed in other transactions (5.1); estimating the premium through propensity score matching to restrict the comparison to highly similar transactions within a given zip code and year (5.2); and implementing the semi-structural estimator of Bajari et al. (2012) that removes bias by introducing a Markov structure for property condition (5.4).

A second source of bias can arise if mortgaged buyers systematically hold higher private valuations of real estate, relative to all-cash buyers, or if they lack the same skill in negotiating the sales price below their private valuation. We address this possibility by using novel data on non-accepted offers to construct an off-equilibrium counterfactual, thus allowing mortgaged buyers to differ in valuation or negotiation skill regardless of whether they succeed in purchasing the home in question (5.3).

These exercises result in a mortgage-cash premium between 7.9% and 16.9% with a median of 11.8%, thereby supporting the baseline estimate's internal validity. Apart from internal validity, we confirm the estimate's external validity in Appendix B by reestimating the premium after including properties without a repeat sale and benchmarking our results against the related literature.

We concisely summarize the various estimates in Table 3 and now discuss them in detail.

5.1 Seller Cash Preference as an Instrumental Variable

We first assess the validity of assumption (5) by using seller characteristics as an instrument for $Mortgaged_{i,t}$, which allows us to estimate the mortgage-cash premium under weaker conditions. Importantly, if assumption (5) is violated because of unobserved property condition, then an instrumental variables approach can only provide unbiased estimates if it uses seller characteristics as instruments, not buyer characteristics. Intuitively, a property in poor condition may attract buyers who typically buy all-cash, but the property's seller was determined in its previous sale when its condition was plausibly different.

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 contemporaneous sale of property i. Thus, this instrument is a " leave-one-out-mean", as commonly used in the labor and development literatures (e.g., Townsend 1994). Letting $Cash\ Share_{s(i,t)}$ denote this instrument, our principal identification assumption becomes

$$\mathbb{E}\left[\, Cash\ Share_{s(i,t)} \times \epsilon_{i,t} \middle|\, \alpha_i, \zeta_{z(i),t}, X_{i,t}\right] = 0, \tag{6}$$

where $s(i,t)$ denotes the seller of property i in month t. Assumption (6) states that a property's price does not depend on the seller's preference for all-cash buyers (i.e., $Cash\ Share_{s(i,t)}$), except through the method of financing. As with assumption (5), the inclusion of property fixed effects sharpens the comparison to two sellers of the same property. Thus, we identify the premium using variation in sellers' preference for cash financing, which, as Section 6 shows theoretically, depends on more fundamental preferences over risk and time.

Columns (1)-(2) of Table 4 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, with an F-statistic well above the cutoffs proposed by Stock and Yogo (2005). In column (2), we follow other papers in the literature (e.g., D'Acunto and Rossi 2020) and evaluate assumption (6) 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 sale price after accounting for the method of financing, $Mortgaged_{i,t}$.

Columns (3)-(4) of Table 4 summarize the second stage results. We estimate a mortgage-cash premium of 15.6% in column (3). In column (4), we control for seller characteristics that plausibly affect the sales price independently of the method of financing, namely existing mortgage debt, concurrently purchasing another home, and having a foreign address. This specification gives a similar estimate of 13.9%, and so $Cash\ Share_{s(i,t)}$ does not simply proxy for these other characteristics.

The similarity of results between Tables 2 and 4 strongly support the baseline results' validity. In fact, the point estimates are so similar 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). Therefore, we conclude that unobserved, time-varying characteristics of the property do not appear to bias the estimated mortgage-cash premium.

5.2 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 4. Then, we match each mortgaged transaction to a counterfactual all-cash transaction within the same zip code and year.

Panels (a) and (b) of Table 5 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 Table 5. 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.

5.3 Non-Accepted Offers as a Counterfactual

We next use data on non-accepted offers to construct a counterfactual sales price under the alternative method of financing. This approach allows us to explicitly control for both the sorting of would-be mortgaged buyers into certain types of properties and the propensity of all-cash buyers to win transactions because of, say, sophistication or negotiating skill.

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

$$\mu \equiv \mathbb{E}[\underbrace{\left(\pi_{i,t}^{M} + \log\left(P_{i,t}^{N}\right)\right)}_{\log\left(P_{i,t}^{M}\right)} - \underbrace{\left(\pi_{i,t}^{C} + \log\left(P_{i,t}^{N}\right)\right)}_{\log\left(P_{i,t}^{C}\right)} |\theta_{i,t}, \mathcal{M}_{i,t}] = \mathbb{E}\left[\pi_{i,t}^{M} - \pi_{i,t}^{C} \middle| \theta_{i,t}, \mathcal{M}_{i,t}\right]. \tag{7}$$

Equation (7) implies that the mortgage-cash premium equals the difference-in-difference between winning versus losing offers between mortgage versus cash-financed purchases. Notably, any price-relevant information contained in non-accepted offer prices (e.g., perceived seller urgency) is differenced out.

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

$$\begin{aligned} \log\left(Price_{i,j,t}\right) = \mu\left(Mortgaged_{i,j,t} \times Winning_{i,j,t}\right) + ... \\ ... + \psi_0 Winning_{i,j,t} + \psi_1 Mortgaged_{i,j,t} + \psi_2 Z_{i,t} + \tau_t + \upsilon_{i,j,t}, \end{aligned} \tag{8}$$

where i, j, and t index property, offer, and month; $Mortgaged_{i,j,t}$ indicates if j is an offer with a positive loan amount; $Winning_{i,j,t}$ indicates if j is the accepted offer; $Price_{i,j,t}$ is the price offered by j; τ_t is a month fixed effect; and the controls in $Z_{i,t}$ are a vector of fixed effects for the number offers on the property and the property's list price.[7]

The identification assumption associated with equation (8) is

$$\mathbb{E}[Mortgaged_{i,j,t} \times Winning_{i,j,t} \times \upsilon_{i,j,t} | \tau_t, Z_{i,t}, Mortgaged_{i,j,t}, Winning_{i,j,t}] = 0. \tag{9}$$

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 self-select into properties in a particular condition. Second, conditioning on $Winning_{i,j,t}$ allows us to identify the premium even if mortgaged and all-cash buyers differ in sophistication, negotiation skill, or other characteristics that make them more likely to win.

Column (1) of Table 6 begins by estimating equation (8) on the set of winning offers,

[7]We cannot include property or zip code-month fixed effects because we do not observe a property identifier and have limited sample size. Controlling for a property's list price leads to a large R-squared and so likely captures the same factors as would otherwise by captured by these fixed effects.

which is the same as estimating equation (4) and, thus, provides a benchmark. The estimated premium of 12.2% lies close to the estimate obtained from the ZTRAX dataset in Table 2, supporting its external validity. The remaining columns include both winning and non-accepted offers, and they imply a mortgage-cash premium between 7.9% and 13.4%. Notably, we obtain an R-squared of 83% after controlling for the list price in column (4). Thus, there is little variation in $\upsilon_{i,j,t}$ to violate assumption (9).

Our offer-level results support both the internal and external validity of the baseline estimates in Table 2, as the offer-level results make use of a separate dataset and a much weaker identification assumption.

5.4 Additional Robustness

We estimate a premium of 14.9% using the semi-structural estimator of Bajari et al. (2012), as we describe in Appendix B.1. In Appendix B.2, we estimate a premium of 16.1% after including properties without a repeat sale. Lastly, in Appendix B.3, we show that our results agree with contemporaneous papers studying the price of cash-financed purchases, after adjusting for differences in sample. Collectively, these findings further support the internal and external validity of the baseline estimate.

6 Calibrated Framework

A positive mortgage-cash premium does not contradict the original Modigliani and Miller (1958) logic as long as standard 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). Then, we describe the calibration results (6.3) and relate them to our empirical findings (6.4).

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 either learning that the transaction has failed, which occurs with probability q, or receiving the offered price, 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. Let r denote the equivalent rate at which consumption in one month is discounted.[8]

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

[8] 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 $e^{\rho} = e^{r(1-\gamma)} - 1$.

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,

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

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.[9] 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. This latter setup would be inappropriate for three reasons. 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).

[9] Equation (13) should be interpreted as an approximate equilibrium condition for two reasons. First, cash-financed purchases can fail and do not close immediately, although they do fail much less often and close more quickly than mortgaged purchases (NAR 2018). Therefore, equation (13) is 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. Second, 14% of sellers have only one offer, based on the offer-level dataset, and not all sellers may have both all-cash and mortgage-financed offers. We assume such sellers are kept to their reservation utility by the possibility that a second offer under the alternative financing method will arrive.

Theoretical Mortgage-Cash Premium

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

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, and two micro-foundations 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. In addition, 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 $1 - e^{-a}$ of the sales price. For ease of language, we shall simply refer to $D_{i,t}$ as the seller's "simultaneous down payment". Lastly, 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.

These extensions imply the following theoretical mortgage-cash premium,

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

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

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

Equation (16) implies that sellers require a higher premium when their own mortgage debt overhang is larger (i.e., ℓ) or when they have a simultaneous down payment (i.e., δ). These fixed obligations raise the mortgage-cash premium insofar as sellers receive a lower price after a failed transaction (i.e., $\kappa > 0$).

In Appendix C, we provide two micro-foundations for the price cut after failure, κ. Both micro-foundations begin with the observation that, in reality, sellers cannot guarantee the arrival of an offer in month $t + 1$. First, we consider an exogenous offer arrival probability, and we suppose the seller continues to cut her list price until, eventually, an offer does arrive. Second, we consider an endogenous arrival probability that, following Guren (2018), increases concavely in the price cut. Accordingly, sellers choose κ to maximize sale probability.

6.3 Calibrated Mortgage-Cash Premium

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 standard 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.

Parameter Values

We parameterize $q = 6.3\%$ based on the share of purchase transactions that were terminated divided by the share of transactions with a mortgaged buyer over 2015-2020, according to a survey of real estate agents conducted by the National Association of Realtors (NAR 2018, 2020). For robustness, we also parameterize $q = 12\%$ based on the average mortgage application denial rate over 1994-2016, based on data from the Home Mortgage Disclosure Act (HMDA). For reasons described in Appendix C, both parameterizations likely constitute an upper bound, since, in the first case, cash-financed purchases can also fail, and, in the second case, 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 $\kappa = 6.0\%$ and $\kappa = 11.1\%$ according to the two micro-foundations described in Appendix C.

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

Calibration Results

Table 8 summarizes the calibrated mortgage-cash premium, μ, across parameterizations. Columns (1)-(5) report the calibrated premium based on the basic framework described in Section 6.1. The results range from 0.7% under our baseline parameterization in column (1) to 1.3% when parameterizing κ using the price cut that maximizes sale probability in column

(4). This latter parameterization constitutes an upper bound, since, in reality, sellers choose κ to trade off a higher sale probability with a lower price, whereas the urgent seller from our framework is concerned only with the former. Indeed, all of the parameterizations of κ used in Table 8 exceed the average value reported by homeowners in our survey from Section 7.2 (4.1%). Lastly, 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 2.6%.

6.4 Relationship between Calibrated and Estimated Premium

We would like to compare the calibrated mortgage-cash premium with the estimated premium, but, in order to do so, we must first evaluate whether the estimated premium varies by the degree of mortgage transaction friction. The presence of such variation jointly supports the validity of our modelling assumptions and the identification assumption (5).

First, we calibrate the mortgage-cash premium for each of the time periods shown in columns (2)-(5) of Table 2. We plot the calibrated and estimated premiums together in Figure 2. The calibrated and estimated premiums covary over time. In particular, 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). Notably, however, the gap between the estimated and calibrated premiums remains stable at between 8% and 10%. Similarly, Appendix Figure A2 shows how the calibrated and estimated premiums covary over the amount of leverage in a transaction, but there is a stable gap between the two.

Next, in Table 9 we reestimate equation (4) after interacting $Mortgaged_{i,t}$ with empirical proxies for the degree of friction in a transaction.[10] The estimates in columns (1)-(2) imply that sellers require a higher premium when they face a large debt overhang (ℓ) and purchase another home in the same month (δ), respectively. Columns (3)-(4) imply that borrowers pay

[10]The proxies are: an indicator for whether the seller's loan-to-value ratio exceeds 50%; an indicator for whether the seller purchases another home in the same month; an indicator for whether the home was the seller's primary residence and will be the buyer's primary residence; an indicator for whether the purchase is financed with a loan insured by the Federal Housing Administration; the loan application denial rate in the associated county and year; and an indicator for whether the average sales price in the associated zip code and month exceeds the $250,000 threshold above which an appraisal is required.

a larger premium when lenders are more likely to reject them (q), measured either by having an FHA-insured loan (e.g., Gete and Reher 2021) or by the county's contemporaneous denial rate. In column (5), we estimate a larger premium in markets where mortgage regulators typically require an appraisal and, thus, higher closing costs (ξ).

Notably, we estimate a positive and significant coefficient on $Mortgaged_{i,t}$ throughout Table 9, not only on its interaction with the proxies for mortgage transaction friction. As with Figure 2, this finding implies that the mortgage-cash premium covaries with the degree of friction, but even mortgaged purchases with relatively little friction still command a significant premium.

Collectively, a standard model with frictions can explain at most 2.6% of the estimated premium of 11.7% from Table 2. Reconciling the remaining 9.1% must appeal to one of two sets of explanations. First, the estimated premium of 11.7% may suffer severe upward bias, in which case there is no puzzle. However, such severe upward bias is unlikely given the stability of estimates across the exhaustive set of exercises described in Section 5. Even the most conservatively estimated premium in Table 3 (7.9%) exceeds the most aggressively calibrated premium in Table 8 (2.6%) by a factor of three. The second set of explanations involve deviating from the rational framework just described.

7 Non-Standard Beliefs, Optimization, and Preferences

We can partition the non-standard explanations of the mortgage-cash premium puzzle into three categories (DellaVigna 2009). First, home sellers may hold different beliefs about the probability of failure (q) and the expected price cut after failure (κ) than those shown in Table 7. Second, sellers may hold the same beliefs as those shown in Table 7, but they calculate expected utility differently than a rational homeowner would. Third, sellers may have standard beliefs and use these beliefs to perform a standard optimization, but the utility function that they optimize differs from the standard one in Section 6.

We use a survey to assess the ability of non-standard beliefs, non-standard optimization, and non-standard preferences to explain the puzzle. 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 discuss evidence of standard (7.3) and non-standard channels (7.3-7.6). Appendix D has details.

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.

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., $\ell e^a = 30\%$). 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., $\delta e^a = 15\%$). To match our framework, this down payment must be made within T weeks, where T is randomized to equal either five or seven. Importantly, we state all quantities in both percent and dollar terms to address issues related to non-proportional thinking.[11]

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

[11] 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."*

the second buyer would pay using mortgage financing. The latter transaction would close in T weeks, which, in keeping with the framework, is intentionally aligned with the seller's deadline from the previous paragraph. The all-cash transaction would close in two weeks, which, as in our framework, implies an average difference in closing periods of four weeks.[12] We also state that the all-cash transaction has "almost no risk" of failing, whereas "there is a chance that the mortgaged buyer will not be able to secure money from their lender". In the latter case, the respondent would need to relist their home in T weeks.

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 two strongest reasons for doing so from a list of four predefined options and a fifth free-response option (i.e., "other").

The second block of questions elicits the respondent's mortgage-cash premium through a sequence of pairwise comparisons, which reduces bias from cognitive overload. 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 5 pps until reaching 20%. 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. For simplicity, we call $Premium_k$ the "observed premium".

In the third block, we elicit the respondent's beliefs about transaction risk, corresponding to the objects q and κ from Section 6. Respondents select a value for q on a sliding scale with a large upper bound of 20% to accommodate non-standard beliefs. Similarly, respondents select the price level at which they would relist their home after a transaction

[12] Our experimental setting features a positive closing period for all-cash offers, which is more realistic than the instantaneous closing period in our framework. We set an all-cash closing period of two weeks, equal to the longest minimum closing period across the following four major iBuyers, based on their company websites: Zillow Offers (7 days), Offerpad (8 days), RedfinNow (10 days), and Opendoor (14 days).

failure, corresponding to $Be^{-\kappa}$. Then, to assess non-proportional thinking, we ask for the equivalent percent price cut. The price level implied by this percent cut lies within 5% of the directly elicited price level for 77% of respondents, suggesting that most have a reasonable degree of numeracy and take the survey seriously.

Finally, we collect information about the respondent's annual household income, age, state of residence, education, experience in real estate markets, and risk aversion. Placing such easy-to-answer questions at the survey's end avoids fatiguing respondents before the more important questions related to the thought experiment.

7.2 Survey Results

Our resulting dataset contains information on 938 U.S. homeowners, summarized in Table 10. All respondents took the survey in April 2021, an unprecedented "seller's market" due to low interest rates, strong demand, and limited supply (Gascon and Haas 2020). Consequently, the mortgage-cash premium elicited through the survey likely provides a conservative estimate of the steady state premium (e.g., low κ, q).

We highlight four takeaways from Table 10. First, the average respondent's mortgage-cash premium of 10.0%, shown in panel (a), lies close to our baseline estimate of 11.7% from Table 2. This value also falls 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.

Second, also shown in panel (a), the average respondent's beliefs about the probability of transaction failure (13.3%) and price cut after failure (4.1%) do not depart dramatically from the set of parameter values used in our calibrations. In particular, these beliefs most closely resemble the parameterization in column (5) in Table 8, in which we parameterize q using the average mortgage denial rate and κ based on the forced sale discount estimated by Campbell, Giglio and Pathak (2011).

Third, turning to panel (c), respondents appeal to a shorter closing period when asked to explain their preference for an all-cash offer, holding price fixed. In particular, 73% cite the following as one of their two most important motivations: *"Even if I was completely certain that the Mortgaged Buyer would follow through, they would still take a longer time to close."* This response holds risk fixed (i.e., "completely certain") to isolate the role of impatience. A smaller share of participants appeal to explanations related to the negative consequences of transaction failure: illiquidity (i.e., "cover my expenses"); regret (i.e., "could have received a higher price"); and low realized return (i.e., "not realize an attractive financial return").

Fourth, panel (b) shows how survey respondents are generally representative of U.S. homeowners, and, if anything, have higher levels of education and income. For example, the share of respondents with a bachelor's degree (71.8%) exceeds the average share among U.S. homeowners (40.1%), per the 2019 American Housing Survey. Moreover, the median respondent's household earns \$3,012 more per year than that of the median U.S. homeowner, per the 2019 American Community Survey. Geographically, Appendix Figure A3 shows how the survey covers all U.S. states except Wyoming in similar proportion to state population.

7.3 Standard Channels

A standard, rational framework such as that from Section 6 would predict that the mortgage-cash premium increases in the seller's expected probability of transaction failure (i.e., q) and price cut after failure (i.e., κ). Column (1) of Table 11 confirms this prediction. Interestingly, comparing the R-squared in columns (1) and (2) implies that variation in beliefs alone explains 15% of the variation in the mortgage-cash premium, while demographic and geographic characteristics only explain an additional 5 pps. Also consistent with a standard model, column (3) shows how respondents require a 0.8 pps larger premium when their stated risk aversion in financial matters is high.

However, such a framework cannot explain the magnitude of respondents' mortgage-cash premium. Figure 3 plots respondents' observed premium against the theoretical premium implied by their beliefs q and κ, based on equation (16). The observed premium lies well above the theoretical premium for almost all respondents, with an average difference of 7 pps. This finding replicates the puzzling discrepancy between the estimated and calibrated values of the mortgage-cash premium from Tables 2 and 8, marked by the red open circle in the figure. We now discuss several non-standard explanations of this puzzle.

7.4 Non-Standard Beliefs

Sellers may hold inflated beliefs about the probability that a mortgaged transaction fails (q). One of the most-well known theories of belief inflation begins with the notion of representativeness (e.g., Kahneman and Tversky 1972; Bordalo, Gennaioli and Shleifer 2012, 2013). According to this logic, failure stands out as a unique (i.e., salient) outcome of accepting a mortgaged offer, and so sellers assign mortgaged transactions an unrealistically high probability of failing. Empirically, however, the average survey respondent's value of q shown in Table 10 lies only 11% (1.3 pps) above the largest parameterization in Table 7, suggesting that belief inflation is not extreme. Moreover, even based on this somewhat large probability, the theoretical premium for most respondents lies well below their observed premium, as shown in Figure 3. Thus, there is little evidence that inflated beliefs can explain the puzzle.

7.5 Non-Standard Optimization

Sellers with standard beliefs and preferences may still require a large mortgage-cash premium if they do not solve equation (13), the optimality condition, as a typical rational seller would. Such non-standard optimization may, for example, lead sellers to perceive a mortgage-cash premium of 3% as "small", whereas it actually translates into a large dollar value. However, in both reality and in our survey, changes in real estate prices are expressed in both dollar and percent terms. Moreover, households tend to focus on dollar values and, consequently, to evaluate large financial investments under inflated percentages (e.g., Shue and Townsend 2021). This behavior actually implies a lower mortgage-cash premium in percent terms. Lastly, 77% of survey respondents exhibit a high degree of numeracy according to the measure in Table 10, and the remaining 23% do not require a significantly higher premium, per column (4) of Table 11. In light of these remarks, non-standard optimization due to non-proportional thinking is an unlikely explanation of the puzzle. More generally, explanations based on non-standard optimization are unlikely to resolve why we continue to find a 7% puzzle among the well-educated homeowners in our survey.

7.6 Non-Standard Preferences

Since neither non-standard beliefs nor non-standard optimization can explain the magnitude of the mortgage-cash premium, we turn to non-standard preferences. We find evidence consistent two non-standard preference structures: present focus and reference dependence

Present Focus

Present focus describes a large class of models in which agents prioritize present flows of experienced utility (Ericson and Laibson 2019). Three pieces of evidence support explanations based on present focus. We have already mentioned the first piece: 73% of respondents cite a shorter closing period as the primary appeal of an all-cash offer, holding risk fixed. This far exceeds the share citing explanations related to reduced downside risk. Second, among the 8% of respondents who explain their preference for all-cash offers by free response, the two most commonly used words are "time" and "weeks", as shown in Appendix Figure A4.

The third piece of evidence supports present bias (i.e., quasi-hyperbolic discounting), a particular form of present focus in which agents discount future utility non-exponentially (e.g., Laibson 1997; O'Donoghue and Rabin 1999; Benhabib, Bisin and Schotter 2010). To structure the discussion, suppose that a seller in month t discounts consumption in month $t + \Delta t$ by the factor $\beta e^{-r\Delta t}$, where $\beta < 1$ governs the degree of present bias. We emphasize the basic point by considering a stylized framework in which mortgaged transactions have no risk of failure (i.e., $q = 0$), but they still take Δt months longer to close. For now, suppose sellers consume most of their sale proceeds. The theoretical mortgage-cash premium then equals

$$\mu = r\Delta t - \log(\beta) . \tag{18}$$

Equation (18) suggests that a plausible degree of present bias can explain the wedge between the empirical and theoretical premium. As a back-of-envelope calibration, our baseline parameterization of $r\Delta t = 0.4\%$ along with a standard present bias factor of $\beta = 0.90$ (e.g., Cohen et al. 2020) implies a mortgage-cash premium of $\mu = 10.9\%$, close to the estimates from both our observational and experimental datasets.[13] More importantly, equation (18) implies that marginally elongating the mortgage closing period will not significantly affect the mortgage-cash premium, insofar as households have a small exponential discount rate over short horizons (i.e., $\frac{\partial \mu}{\partial \Delta t} = r \approx 0$). Indeed, we find that randomly elongating the mortgage closing period by two weeks has no significant effect on a survey respondent's observed premium, as shown in column (1) of Table 12. This finding suggests that the discount factor falls steeply after the closing of the all-cash offer, and then it quickly flattens thereafter.

A proper framework of present-biased home sellers would require two additional features. First, such a framework would require some utility loss from, say, psychological stress incurred per Δt units of time. This feature accounts for the fact that present bias concerns impatience over experienced utility, as opposed to impatience over monetary rewards (e.g., O'Donoghue and Rabin 2015). It also realistically captures the high stress levels that homeowners typically incur with selling their home. Second, in our setting, the shortest period Δt during which a deal can close is one to two weeks, per footnote 12. This differs from laboratory settings, in which Δt can be as small as a few hours (Ericson and Laibson 2019). Thus, our assumption that the "later" period sets in after at least a week aligns with quasi-hyperbolic consumption-savings models, in which present bias affects investment over periods of months (e.g., Angeletos et al. 2001). More broadly, our evidence suggests that present bias helps explain large financial decisions, not only decisions made in the lab.

[13] We emphasize that our survey does not allow for a rigorous elicitation of present bias because it lacks important features such as convex budget sets (e.g., Andreoni and Sprenger 2012). When designing the survey, we had in mind the more basic goal of assessing the relative importance of non-standard beliefs, optimization, and preferences.

Reference Dependence

Sellers with reference dependent preferences experience additional utility loss when they receive a sale price less than the price around which they have formed a reference point. It is well-known that such reference dependence affects the sale price in real estate transactions (e.g., Genesove and Mayer 2001). However, it is theoretically unclear whether and how reference dependence affects the relative price of a transaction's method of financing (i.e., mortgage-cash premium). In particular, a reference point can only affect the mortgage-cash premium if it lies within the support of potential sale prices: $\left\{P_{i,t}^{R}, P_{i,t}^{C}, P_{i,t}^{M}\right\}$. Otherwise, the two methods of financing come with the same probability of loss, and so loss aversion has no net effect on the mortgage-cash premium.

We clarify this point by applying the setup from Kőszegi and Rabin (2006) to our setting. Suppose home sellers have gain-loss utility that enters additively into the standard utility function, with coefficient $\eta > 0$. Let $\overline{P}_{i,t} \equiv P_{i,t}^{C} e^{-\Lambda}$ denote the reference point, which sellers value according to the isoelastic utility function in equation (10). Gains relative to this reference point, in utils, add one-for-one to the gain-loss component of utility. Losses, by contrast, subtract a multiple $\lambda > 1$. We emphasize the key intuition by supposing $r = 0$, a simplification that is symmetric to that just made in the case of present bias.

We consider the case in which the reference point lies between the prices associated with transaction failure and cash financing: $\kappa > \Lambda > 0$. Intuitively, this case features loss if and only if the seller accepts a mortgaged offer that subsequently fails, as we formalize in Proposition C.1 of the online appendix. Otherwise, the seller experiences loss under both methods of financing (i.e., $\Lambda > \kappa$), in which case loss aversion does not affect their relative price. Or, the seller experiences loss under cash financing (i.e., $0 > \Lambda$), in which case loss aversion could require a lower mortgage-cash premium.

The mortgage-cash premium then equals

$$\mu = \frac{1}{\gamma - 1}\left[q\left(e^{\kappa(\gamma-1)} - 1\right)\right] + q\frac{\eta}{1+\eta}\left(\lambda - 1\right)\left(\frac{e^{-\Lambda(1-\gamma)}}{1-\gamma} - \frac{e^{-\kappa(1-\gamma)}}{1-\gamma}\right), \tag{19}$$

where, again, this equation follows from applying Kőszegi and Rabin (2006) to our setting.

Intuitively, when sellers have gain-loss utility ($\eta > 0$) and exhibit loss aversion ($\lambda > 1$), they require mortgaged buyers to pay a higher premium because failed transactions constitute a loss relative to their reference point ($\kappa > \Lambda$).

The first term in equation (19) only reflects standard channels, so that loss aversion additively increases the mortgage-cash premium through the second term. Consider a back-of-envelope calculation in which the reference point equals the cash-financed offer price (i.e., $\Lambda = 0$) and sellers have purely gain-loss utility (i.e., $\frac{\eta}{1+\eta} = 1$). Then, with a plausible loss aversion of $\lambda = 2.25$ (e.g., Tversky and Kahneman 1992) and the parameterization in column (5) of Table 8, the second term in equation (19) equals 0.9%. This calculation suggests that loss aversion can reduce the 8 pps mortgage-cash premium puzzle by 11% (0.9 pps).

Quantitatively, therefore, it is not clear that loss aversion alone can resolve the puzzle.[14] Qualitatively, though, two pieces of evidence suggest that the mortgage-cash premium partly reflects reference dependence around non-accepted offer prices. First, 44% of survey respondents cite post-failure "regret that I could have received a higher price from the cash buyer" as one of the two most important drawbacks of a mortgaged offer, as shown in Table 10. Second, column (2) of Table 12 uses the offer-level methodology from Section 5.3 to show how sellers with more offers and, thus, a higher reference point, require a higher mortgage-cash premium. Notably, the number of offers is unlikely to simply proxy for market liquidity, since liquid markets make transaction failure less costly, thus implying a lower premium.

Our collective takeaway from the survey evidence is that incorporating non-standard preferences better explains the mortgage-cash premium puzzle than either non-standard beliefs or non-standard optimization. We discuss the accompanying policy implications in our conclusion.

[14] A similar caveat as in footnote 13 applies, in that a proper treatment of reference dependence requires taking a stance on more parameters than we can elicit from a single survey. For example, Andersen et al. (2021) do so through a structural estimation.

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 realistically calibrated model of rational home sellers implies a premium of only 3%. Our survey points to non-standard preferences 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 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 behavioral channels that amplify them. The former route has an outsized impact because of behavioral amplification. The latter route requires identifying the form of behavioral preferences that best explains the mortgage-cash premium. 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* .

Andreoni, J. and Sprenger, C.: 2012, Estimating Time Preferences from Convex Budgets, *American Economic Review* .

Angeletos, G.-M., Laibson, D., Repetto, A., Tobacman, J. and Weinberg, S.: 2001, The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation, *Journal of Economic Perspectives* .

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* .

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* .

Benhabib, J., Bisin, A. and Schotter, A.: 2010, Present-Bias, Quasi-Hyperbolic Discounting, and Fixed Costs, *Games and Economic Behavior* .

Bordalo, P., Gennaioli, N. and Shleifer, A.: 2012, Salience Theory of Choice under Risk, *The Quarterly Journal of Economics* .

Bordalo, P., Gennaioli, N. and Shleifer, A.: 2013, Salience and Consumer Choice, *Journal of Political Economy* .

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

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

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* .

Cohen, J., Ericson, K. M., Laibson, D. and White, J. M.: 2020, Measuring Time Preferences, *Journal of Economic Literature* .

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* .

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

Epper, T., Fehr, E., Fehr-Duda, H., T, K. C., Lassen, D. D., Leth-Petersen, S. and Rasmussen, G. N.: 2020, Time Discounting and Wealth Inequality, *American Economic Review* .

Ericson, K. M. and Laibson, D.: 2019, Intertemporal Choice, *Handbook of Behavioral Economics - Foundations and Applications*.

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* .

Frederick, S., Loewenstein, G. and O'Donoghue, T.: 2002, Time Discounting and Time Preference: A Critical Review, *Journal of Economic Literature* .

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.: 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* .

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.

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.

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* .

Kahneman, D. and Tversky, A.: 1972, Subjective Probability: A Judgment of Representativeness, *Cognitive Psychology* .

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* .

Laibson, D., Maxted, P. and Moll, B.: 2020, Present Bias Amplifies the Household Balance-Sheet Channels of Macroeconomic Policy.

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* .

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

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* .

O'Donoghue, T. and Rabin, M.: 2015, Present Bias: Lessons Learned and To Be Learned, *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.: 2016, Housing and Macroeconomics, *Handbook of Macroeconomics*.

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* .

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

Shue, K. and Townsend, R. R.: 2021, Can the Market Multiply and Divide? Non-Proportional Thinking in Financial Markets, *The Journal of Finance* .

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.

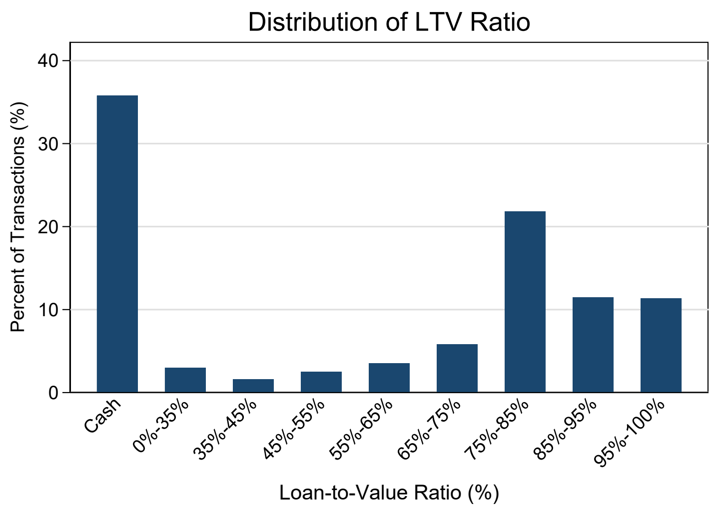
Zillow: 2018, Consumer Housing Trends Report, *Technical report.*

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

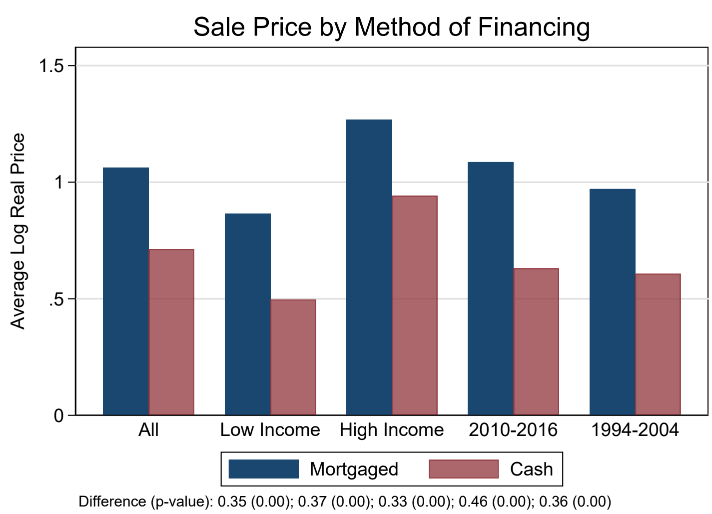
Zillow: 2021, The Home-Selling Process for an All-Cash Deal, *Technical report.*

Figures and Tables

Figure 1: Facts About All-Cash Purchases



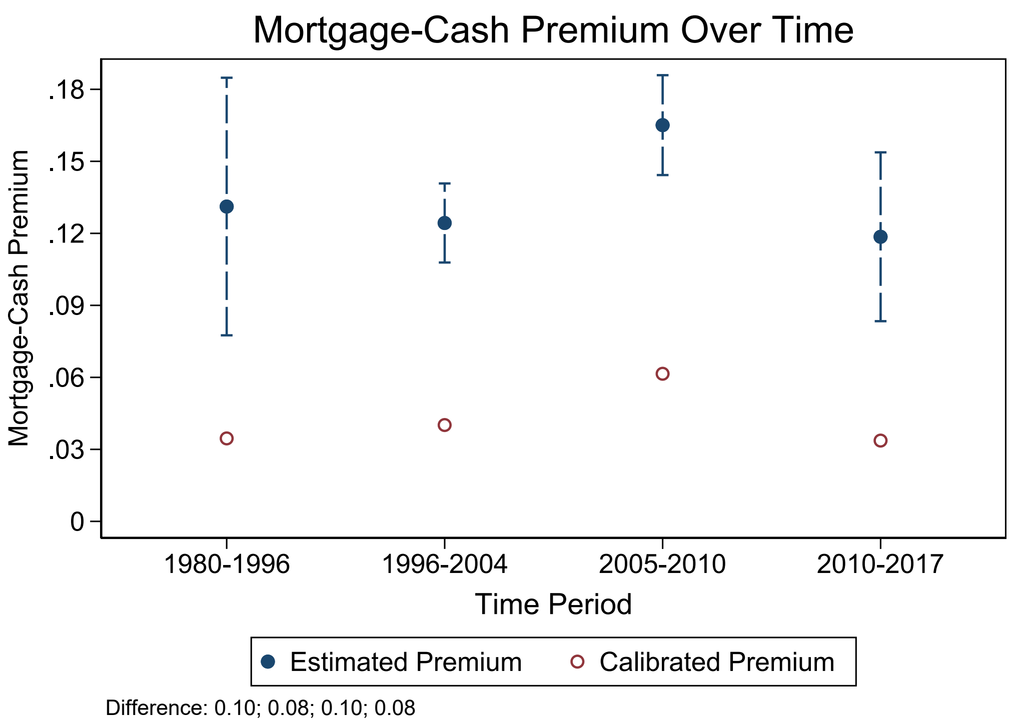
(a) Distribution of Loan-to-Value Ratio



(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,

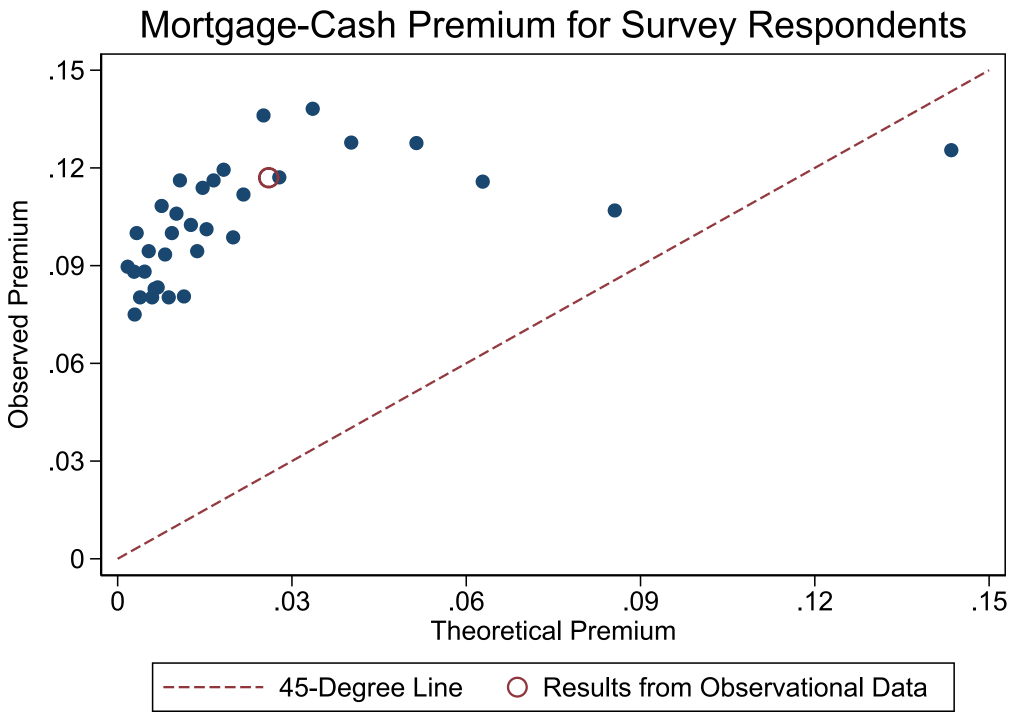
$$\log\left(Price_{i,t}\right) = \mu Mortgaged_{i,t} + \psi X_{i,t} + \zeta_{z(i),t} + \alpha_i + \epsilon_{i,t}, \quad t \in \mathcal{T}$$

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 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: Observed and Theoretical 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

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

where the parameters are defined in the note to Table 8. The figure parameterizes q and κ according to the percent of the time that the respondent believes a mortgage-financed offer will fail and the log price cut that the respondent would impose after a transaction failure, respectively. The preference parameters r and γ are parameterized as in Table 7, where r has been amortized to match the difference between the mortgaged offer's closing period and the all-cash offer's closing period. The remaining mortgage market parameters are parameterized according to the survey's thought experiment: $\xi = 0$; $\ell e^{\alpha} = 0.30$; $\delta e^{\alpha} = 0.15$; and $a = 0.06$. 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.

Table 1: Summary Statistics

	Observations	Mean	Standard Deviation
(a) Transaction-Level Dataset:			
$Real\ Price_{i,t}$	3,528,981	\$416,813	\$776,315
$Mortgaged_{i,t}$	3,528,981	0.642	0.479
Age_i	2,525,425	28.431	6.682
$Rooms_i$	2,525,437	1.409	1.492
$Bathrooms_i$	2,5254,37	0.237	0.715
$Stories_i$	2,525,425	1.096	0.111
$Air\ Conditioning_i$	2,525,438	0.217	0.239
$Detached_i$	3,528,981	0.405	0.491
$High\ Seller\ LTV_{i,t}$	3,528,981	0.068	0.252
$Same\text{-}Month\ Purchase_{i,t}$	3,528,981	0.015	0.122
$FHA\ Borrower_{i,t}$	3,528,981	0.196	0.397
$Standardized\ Denial\ Rate_{c(i),t}$	3,528,981	0.009	0.569
$Appraisal\ Required_{z(i),t}$	3,068,605	0.517	0.500
$Cash\ Share_{s(i,t)}$	3,528,981	0.179	0.327
(b) Offer-Level Dataset:			
$Real\ Price_{i,j,t}$	6,427	\$512,662	\$359,480
$Mortgaged_{i,j,t}$	6,431	0.609	0.488
$Winning_{i,j,t}$	6,431	0.166	0.372
Baseline Number of Transactions: 426,256			
Baseline Number of Offers: 5,963			

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 $Single\ Family_i$ indicate if i has air conditioning and is a detached single-family home, respectively; $High\ Seller\ LTV_{i,t}$ 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; $FHA\ Borrower_{i,t}$ indicates if the purchase is financed with a loan insured by the Federal Housing Administration; $Standardized\ Denial\ Rate_{c(i),t}$ is the loan application denial rate in the associated county, $c(i)$, and year, normalized to have a mean of zero and unit variance and then assigned a value of zero when unobserved; $Appraisal\ Required_{z(i),t}$ is an indicator for whether the average sales price in the associated zip code $z(i)$ and month exceeds the \$250,000 threshold above which an appraisal is required; and $Cash\ Share_{s(i,t)}$ 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. 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 May 2020 through October 2020. 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; and $Real\ Price_{i,j,t}$ is the price offered by j in 2010 dollars. The bottom rows show the number of observations from each dataset that can be used in Tables 2 and 6, 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)
$Mortgaged_{i,t}$	0.191	0.117	0.131	0.124	0.165	0.119	0.090	0.088	0.112	0.108
	(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 Institutional	Non Foreign
Zip Code-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hedonic-Month FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Property FE	No	No	Yes	Yes	Yes	Yes	No	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
Number of Observations	426,256	426,256	27,087	122,683	69,307	20,792	6,651	186,570	333,288	323,956

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 Institutional). Column (9) restricts the sample to properties in which neither the buyer nor seller has an address outside the U.S. (Non Foreign). Standard errors are clustered by property. Data are from the ZTRAX dataset.

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

Outcome:	$\log(Price_{i,t})$	
	Estimated μ	Table
Repeat-Sales-Hedonic, Baseline	0.117 (0.000)	Table 2, Column 1
Repeat-Sales-Hedonic, 1980-1996	0.131 (0.000)	Table 2, Column 2
Repeat-Sales-Hedonic, 1996-2004	0.124 (0.000)	Table 2, Column 3
Repeat-Sales-Hedonic, 2005-2010	0.165 (0.000)	Table 2, Column 4
Repeat-Sales-Hedonic, 2010-2017	0.119 (0.000)	Table 2, Column 5
Repeat-Sales-Hedonic, New Homes	0.090 (0.004)	Table 2, Column 6
Repeat-Sales-Hedonic, No Flips	0.088 (0.000)	Table 2, Column 7
Repeat-Sales-Hedonic, Non Institutional	0.112 (0.000)	Table 2, Column 8
Repeat-Sales-Hedonic, Non Foreign	0.108 (0.000)	Table 2, Column 9
Instrumental Variable	0.139 (0.017)	Table 4, Column 4
Matching	0.169 (0.000)	Table 5, Column 3
Non-Accepted Offers	0.079 (0.002)	Table 6, Column 4
Homeowner Survey	0.100 (0.000)	Table 10, Column 2
Hedonic with Non-Repeat Sales	0.161 (0.000)	Appendix Table A2, Column 1
Semi-Structural (Bajari et al. 2012)	0.149 (0.000)	Appendix Table A3, Column 2
Repeat-Sales-Hedonic, 10%-60% LTV	0.117 (0.001)	Appendix Figure A2
Repeat-Sales-Hedonic, 61%-80% LTV	0.089 (0.000)	Appendix Figure A2
Repeat-Sales-Hedonic, 81%-90% LTV	0.113 (0.000)	Appendix Figure A2
Repeat-Sales-Hedonic, 91%-125% LTV	0.143 (0.000)	Appendix Figure A2

Note: P-values are in parentheses. 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 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 (6). 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 $Mortgaged_{i,t}$. The seller controls in column (4) are $High\ Seller\ LTV_{i,t}$ and $Same\text{-}Month\ Purchase_{i,t}$, defined in the note to Table 1, and an indicator for whether the seller has a foreign address. 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 in the repeat-sales-hedonic specification with no instrument. The remaining notes are the same as in Table 2.

Table 5: 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)
(a) 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)
(b) 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 4 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 6: Robustness to Using Data on Non-Accepted Offers

Outcome:	$\log\left(Price_{i,j,t}\right)$			
	(1)	(2)	(3)	(4)
$Mortgaged_{i,j,t} \times Winning_{i,j,t}$	0.122	0.134	0.128	0.079
	(0.011)	(0.010)	(0.016)	(0.002)
Subsample of Offers	Winning	All	All	All
Month FE	Yes	Yes	Yes	Yes
Mortgaged Control	No	Yes	Yes	Yes
Winning Control	No	Yes	Yes	Yes
Offers-on-Property FE	No	No	Yes	Yes
List Price Control	No	No	No	Yes
R-squared	0.018	0.019	0.033	0.825
Number of Observations	1,005	5,963	5,963	5,963

Note: P-values are in parentheses. This table estimates equation (8), 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

$$\log\left(Price_{i,j,t}\right) = \mu\left(Mortgaged_{i,j,t} \times Winning_{i,j,t}\right) + \psi_0 Winning_{i,j,t} + \psi_1 Mortgaged_{i,j,t} + \psi_2 Z_{i,t} + \tau_t + \upsilon_{i,j,t},$$

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; and $Price_{i,j,t}$ is the price offered by j. The sample in column (1) is restricted to winning offers, which facilitates comparison with our baseline results in Table 2. The sample in the remaining columns includes all offers. Columns (3)-(4) include a vector of fixed effects for the number of offers on the property. Column (4) controls for the property's list price. The observation window is May 2020 through October 2020. Observations are weighted by the number offers on the property. Standard errors are clustered by offer. Data are from the offer-level dataset. 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)	0.063	NAR
	0.12	HMDA
Price Cut after Failure (κ)	0.052	CGP
	0.057	HRY
	0.060	GG
	0.111	Guren
Seller Closing Costs (ξ)	0.013	HUD
Current Loan-to-Value Ratio (ℓe^{-a})	0.331	ZTX
Simultaneous Down Payment (δe^{-a})	0.175	ZTX
(b) Preference Parameters:		
Discount Rate (r), Annualized	0.03	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 divided by the share of transactions with a mortgaged buyer over 2015-2020, from the National Association of Realtors (2018, 2020); HMDA refers to the average mortgage denial rate over 1994-2016, based on data from HMDA; 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; Guren refers to the log price cut relative to the market average that maximizes the probability of sale, based on the average estimates of the quadratic specification in Table 2 of Guren (2018) 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); and 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). Real estate agent fees, denoted a 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	0.013	0.011	0.020	0.022	0.026
<u>Mortgaged Transaction Parameters:</u>								
Probability of Transaction Failure (q)	0.063 (NAR)	0.063 (NAR)	0.063 (NAR)	0.063 (NAR)	0.12 (HMDA)	0.063 (NAR)	0.063 (NAR)	0.063 (NAR)
Price Cut after Failure (κ)	0.052 (CGP)	0.057 (HRY)	0.060 (GG)	0.111 (Guren)	0.052 (CGP)	0.052 (CGP)	0.052 (CGP)	0.052 (CGP)
Seller Closing Costs (ξ)						0.013 (HUD)	0.013 (HUD)	0.013 (HUD)
Current Loan-to-Value Ratio (ℓe^{-a})							0.331 (ZTX)	0.331 (ZTX)
Simultaneous Down Payment (δe^{-a})								0.175 (ZTX)

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^{-a}); 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^{-a}); the seller's coefficient of relative risk aversion (γ); and real estate agent fees paid by the seller (a). Explicitly,

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

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:	$\log\left(Price_{i,t}\right)$				
	(1)	(2)	(3)	(4)	(5)
$Mortgaged_{i,t}$	0.107	0.116	0.116	0.131	0.115
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$Mortgaged_{i,t} \times High\ Seller\ LTV_{i,t}$	0.033				
	(0.000)				
$Mortgaged_{i,t} \times Same\text{-}Month\ Purchase_{i,t}$		0.041			
		(0.032)			
$Mortgaged_{i,t} \times FHA\ Borrower_{i,t}$			0.062		
			(0.001)		
$Mortgaged_{i,t} \times Standardized\ Denial\ Rate_{c(i),t}$				0.024	
				(0.000)	
$Mortgaged_{i,t} \times Appraisal\ Required_{z(i),t}$					0.041
					(0.000)
Interaction as Control	Yes	Yes	Yes	Yes	Yes
Zip Code-Month FE	Yes	Yes	Yes	No	No
Hedonic-Month FE	Yes	Yes	Yes	Yes	Yes
Property FE	Yes	Yes	Yes	Yes	Yes
R-squared	0.907	0.907	0.907	0.828	0.828
Number of Observations	426,256	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 and the costs of these frictions for sellers. Subscripts i and t index property and month. The regression equation is similar to that in Table 2 after interacting $Mortgaged_{i,t}$ with a variable related to transaction frictions: an indicator for whether the seller's loan-to-value ratio is above 50% ($High\ Seller\ LTV_{i,t}$); an indicator for whether the seller purchases another home in the same month ($Same\text{-}Month\ Purchase_{i,t}$); an indicator for whether the purchase is financed with a loan insured by the Federal Housing Administration ($FHA\ Borrower_{i,t}$), which have higher loan application denial rates; the loan application denial rate in the associated county, $c(i)$, and year, normalized to have a mean of zero and unit variance ($Standardized\ Denial\ Rate_{c(i),t}$); and an indicator for whether the average sales price in the associated zip code $z(i)$ and month exceeds the $250,000 threshold above which an appraisal is required ($Appraisal\ Required_{z(i),t}$). The numerator of the loan-to-value ratio used to construct $High\ Seller\ LTV_{i,t}$ 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 to remove mechanical correlation with $Price_{i,t}$. All specifications control separately for the variable interacted with $Mortgaged_{i,t}$. Additional details on each variable's definition are in the note to Table 1. The remaining notes are the same as in Table 2.

Table 10: Experimental Mortgage-Cash Premium and Other Characteristics of Survey Respondents

	Observations	Mean	Standard Deviation
(a) Mortgage-Cash Premium and Beliefs:			
Premium$_k$ (μ)	937	0.100	0.062
Expected Failure Probability$_k$ (q)	938	0.133	0.062
Expected Price Cut$_k$ (κ)	934	0.041	0.061
(b) Background Information:			
Household Income$_k$	938	\$102,202	\$81,916
Age$_k$	938	40.439	12.375
College Educated$_k$	937	0.718	0.45
High Risk Aversion$_k$	938	0.652	0.476
High Numeracy$_k$	938	0.769	0.422
(c) "Why would you prefer to sell your home to the Cash Buyer rather than the Mortgaged Buyer?":			
"If the Mortgaged Buyer backs out and I relist at a lower price, I may not be able to cover my expenses."	919	0.371	0.483
"If the Mortgaged Buyer backs out and I relist at a lower price, I'd regret that I could have received a higher price from the Cash Buyer."	919	0.444	0.497
"If the Mortgaged Buyer backs out and I relist at a lower price, I might not realize an attractive financial return on my home."	919	0.169	0.375
"Even if I was completely certain that the Mortgaged Buyer would follow through, they would still take a longer time to close."	919	0.731	0.444

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. Panel (a) summarizes variables that correspond to parameters from the calibration in Table 8: *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 to account for non-proportional thinking; *Expected Failure Probability*$_k$ is the percent of the time that k believes a mortgaged transaction will fail; and *Expected Price Cut*$_k$ is the log price cut k would impose after a transaction failure. Panel (b) summarizes background variables: *High Risk Aversion*$_k$ indicates if k is "usually reluctant to take risks but sometimes willing to do so" or "never willing to take risks", as opposed to "usually willing to take risks but sometimes reluctant to do so" or "always willing to take risks"; and *High Numeracy*$_k$ indicates if k's response to a question phrased in terms of dollars is within 5% of the same question phrased in terms of percent. Panel (c) summarizes responses to the question of why a respondent would prefer an all-cash offer. Respondents are allowed to select two motives, and so the means do not sum to one across motives. Section 7 and Appendix D contain additional details.

Table 11: Variation in the Experimental Premium by Beliefs

Outcome:	*Premium$_k$*			
	(1)	(2)	(3)	(4)
Expected Failure Probability$_k$	0.361	0.348	0.347	0.341
	(0.000)	(0.000)	(0.000)	(0.000)
Expected Price Cut$_k$	0.121	0.093	0.095	0.078
	(0.002)	(0.021)	(0.017)	(0.059)
High Risk Aversion$_k$			0.008	
			(0.047)	
High Numeracy$_k$				-0.007
				(0.153)
Household Controls	No	Yes	Yes	Yes
State FE	No	Yes	Yes	Yes
R-squared	0.152	0.206	0.209	0.208
Number of Observations	933	930	930	930

Note: P-values are in parentheses. This table estimates regressions of a survey respondent's mortgage-cash premium on her beliefs about transaction risk. Subscript k indexes survey respondent. The regression equation is of the form

$$Premium_k = \psi_0 \left(Expected\ Failure\ Probability_k\right) + \psi_1 \left(Expected\ Price\ Cut_k\right) + \psi X_k + \zeta_{S(i)} + u_i,$$

where observations are U.S. homeowners who partake in the survey; $\zeta_{S(i)}$ is a fixed effect for the survey respondent's state of residence; and X_k is a vector of controls describing the respondent: age; log of household income; and an indicator for whether the respondent has a bachelor's degree. Columns (3)-(4) include additional indicator variables related to risk aversion and numeracy. Standard errors are clustered by survey respondent. The remaining notes are the same as in Table 10.

Table 12: Mortgage-Cash Premium and Non-Standard Preferences

Outcome:	$Premium_k$	$\log\left(Price_{i,j,t}\right)$
	(1)	(2)
(a) Present Bias:		
$Mortgage\ Closing\ Period_k$	-0.001	
	(0.512)	
(b) Reference Dependence:		
$Mortgaged_{i,j,t} \times Winning_{i,j,t}$		0.081
		(0.002)
$Mortgaged_{i,j,t} \times Winning_{i,j,t} \times Offers\text{-}on\text{-}Property_{i,t}$		0.013
		(0.037)
Dataset	Survey	Offer-Level
Other Specification Notes	Table 11 Column (2)	Table 6 Column (4)
R-squared	0.206	0.825
Number of Observations	933	5,963

Note: P-values are in parentheses. This table estimates variants of the specifications in Tables 6 and 11, which assess the consistency of the mortgage-cash premium with non-standard preferences. Subscripts k, i, j, and t index survey respondent, property, offer, and month. Column (1) tests for present bias, which predicts that a marginal lengthening of the mortgage closing period does not affect the mortgage-cash premium. The specification is the same as in column (2) of Table 11, after also controlling for the length of the randomly assigned mortgage closing period for respondent k, in weeks ($Mortgage\ Closing\ Period_k$). Column (2) tests for reference dependence around non-accepted offers, which predicts that sellers with more attractive non-accepted offers require a higher mortgage-cash premium. The specification is the same as in column (4) of Table 6, except that the treatment variable, $Mortgaged_{i,j,t} \times Winning_{i,j,t}$, is interacted with the number of offers for the purchase of property i in month t, normalized to have a mean of zero and unit variance ($Offers\text{-}on\text{-}Property_{i,t}$). Note that the fixed effects for the number of offers on the property absorb the direct effect of this variable. The remaining notes for each column are the same as in the table shown in the row Other Specification Notes.

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. 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. For both types of parties, we observe the party's name, address, and indicators for whether the party has a foreign address or is institutional. Zillow defines institutions somewhat loosely, as it includes trusts in its definition. 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.

The third set of variables are hedonic characteristics, which come from the ZAsmt component of ZTRAX. We catalog these characteristics in Appendix A.4. 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 Offer-Level Dataset

Our secondary dataset comes from a large U.S. real estate brokerage, which we use in the robustness exercise from Section 5.3. The offer-level 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 the brokerage. We collected the data from the brokerage's website in October 2020 using standard automated web browsing methods.[15] Since the brokerage only retains recent offers on its website, the resulting dataset spans May 2020 through October 2020. In terms of coverage, the brokerage dataset covers 53% of U.S. counties on a population-weighted basis.

The unit of observation in the offer-level dataset is an offer for the purchase of a one-to-four family home, which includes condominiums. Out of respect for their clients' privacy, the brokerage reports only the zip code of the property in question, but we do not observe the property's address. We observe the month when the offer was made; the list price; the offer price, expressed as a percent above or below list price; the loan-to-value (LTV) ratio, rounded the nearest 10% for mortgage-financed purchases; and the number of competing offers on the property.

A.3 Additional Datasets

We rely on several additional datasets. One of these datasets contains experimental data from a survey of 938 U.S. homeowners, which we describe in detail in Appendix D. In addition, 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. Lastly, we use data on mortgage denial rates from the Home Mortgage Disclosure Act (HMDA) to calculate the variable $\textit{Standardized Denial Rate}_{c(i),t}$, defined below. The HMDA data that we use are at the county-by-year level and were provided to us by Recursion Co.

[15] In detail, we collect the data from the brokerage's public website using a Google Chrome web driver and the Python module *selenium*. Within the web page for each U.S. zip code covered by the brokerage, the algorithm locates the Offer Insights table within the Market Insights tab. Each observation in this table is an offer made on a property within the zip code. The algorithm extracts the following information: the month when the offer was made; the list price; the offer price, expressed as a percent above or below list price; the loan-to-value ratio; and the number of competing offers on the property. Our data come from running the algorithm in August 2020 and October 2020. The resulting dataset spans May 2020 through October 2020 because offers older than three months are removed from the website.

A.4 Catalog of Variables

ZTRAX Dataset

- $\textit{Mortgaged}_{i,t}$: This variable indicates if the loan amount associated with the purchase of property i in month t is positive.
- $\textit{Price}_{i,t}$: This variable is the sales price associated with the purchase of property i in month t.
- $\textit{Age}_i$: This variable is the number of years from when property i was built.
- $\textit{Rooms}_i$: This variable is the number of overall rooms in property i.
- $\textit{Bathrooms}_i$: This variable is the number of overall bathrooms in property i.
- $\textit{Stories}_i$: This variable is the number of overall stories in property i.
- $\textit{Air Conditioning}_i$: This variable indicates if property i has air conditioning.
- $\textit{Detached}_i$: This variable indicates if property i is a detached single-family home
- $\textit{High Seller LTV}_{i,t}$: 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 $\textit{Price}_{i,t}$. We assign a value of zero to this indicator variable when the raw variable is unobserved.
- $\textit{Same-Month Purchase}_{i,t}$: 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.
- $\textit{FHA Borrower}_{i,t}$: This variable indicates if the purchase of property i in month t is financed with a loan insured by the Federal Housing Administration, based on the indicator variable used by Zillow to flag loan type. We assign a value of zero to this indicator variable when the raw variable is unobserved.
- $\textit{Standardized Denial Rate}_{c(i),t}$: This variable is the loan application denial rate in the county, $c(i)$, and year associated with the purchase of property i in month t. We normalize the raw denial rate to have a mean of zero and unit variance and then assign it a value of zero when it is unobserved. The raw data come from HMDA via Recursion Co and are at the county-year level.
- $\textit{Appraisal Required}_{z(i),t}$: This variable is an indicator for whether the average sales price in the zip code, $z(i)$, and month associated with the purchase of property i in month t exceeds \$250,000. The FDIC requires mortgage-financed purchases with a sales price greater than \$250,000 to have an appraisal. This threshold changed in 2019, after the end of our sample period. 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*$_{i,t}$: 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.

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. For mortgage-financed offers, the observed LTV ratio is rounded to the nearest 10%.

- *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*$_{i,j,t}$.

Survey Dataset

- *Premium*$_k$: This variable equals the midpoint between the minimum premium at which respondent k would prefer the mortgaged offer in our thought experiment and the maximum premium at which k would prefer the all-cash offer. As described in Appendix D, the resulting support for this variable is: $\{0\%, 2.5\%, 7.5\%, 12.5\%, 17.5\%, 22.5\%\}$.

- *Expected Failure Probability*$_k$: This variable is the percent of the time that respondent k believes a mortgage-financed offer will fail.

- *Expected Price Cut*$_k$: This variable is the log price cut respondent k would impose after a transaction failure.

- *Household Income*$_k$: This variable is the annual household income for respondent k.

- *Age*$_k$: This variable is the age of survey respondent k.

- *College Educated*$_k$: This variable indicates if respondent k has a bachelor's degree.

- *High Risk Aversion*$_k$: This variable indicates if, in financial matters, k is "usually reluctant to take risks but sometimes willing to do so" or "never willing to take risks", as opposed to "usually willing to take risks but sometimes reluctant to do so" or "always willing to take risks".

- *High Numeracy*$_k$: This variable indicates if respondent k's answer to the question of how large of a dollar price cut k would impose after a transaction failure lies within 5% of k's answer to the same question phrased in percentage terms.

- *Mortgage Closing Period*$_k$: This variable equals the number of weeks that a mortgaged transaction takes to close in the thought experiment to which k was randomly assigned. The support is: $\{5, 7\}$.

B Additional Robustness

We perform additional robustness exercises referenced in Section 5.4.

B.1 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 (9). 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 $\log\left(Price_{i,t}\right)$ and $Mortgaged_{i,t}$ against the property and zip code-month fixed effects in our repeat-sales methodology.[16] 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

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

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

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

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.

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

Finally, suppose market participants rely on rational expectations such that

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

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

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

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

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

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

Appendix Table A3 reports the results. We estimate a mortgage-cash premium of 10.9% using this semi-structural approach, per the result in column (1). 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.[17] This results in a similar mortgage-cash premium of 14.9%, shown in column (2).

Along with the instrumental variable results from Section 5.1 and the results based on non-accepted offers from Section 5.3, 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.

[17]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., $\varphi_t^P\left(n\right)$, $\varphi_t^M\left(n\right)$) 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 $\log\left(Price_{i,t}\right)$ and $Mortgaged_{i,t}$ against property fixed effects, means that there is not enough variation to estimate $\varphi_t^P\left(n\right)$ and $\varphi_t^M\left(n\right)$.

B.2 Properties without a Repeat Sale

When expanding the sample to include properties without a repeat sale, we obtain a larger point estimate of 16.1% and a lower R-squared of 58%, as shown in Appendix Table A2. 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.

B.3 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 A4, 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.[18] Since mortgage-financed purchases account for 17% of all purchases in the sample studied by Buchak et al. (2020), as shown in Appendix Table A4, 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 A4, 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.

[18]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: $P^M_{i,t} = e^{\mu} P^C_{i,t} = e^{\mu} P^{\mathcal{I}}_{i,t}$. Then

$$\iota = \mathbb{E}\left[w \log\left(P^M_{i,t}\right) + (1-w)\log\left(P^C_{i,t}\right)\right] - \mathbb{E}\left[\log\left(P^{\mathcal{I}}_{i,t}\right)\right] = w\mu.$$

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 the expressions for the theoretical mortgage-cash premium with the non-standard preferences described in Section 7.6.

C.1 Derivation of the Theoretical Mortgage-Cash Premium

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

$$\frac{\left(P_{i,t}^{C}e^{-a}-L_{i,t}-D_{i,t}\right)^{1-\gamma}}{1-\gamma}=e^{-\rho}\left[\frac{\left(P_{i,t}^{M}e^{-a-\xi}-L_{i,t}-D_{i,t}\right)^{1-\gamma}}{(1-q)^{-1}(1-\gamma)}+\frac{\left(P_{i,t+1}^{R}e^{-a}-L_{i,t}-D_{i,t}\right)^{1-\gamma}}{q^{-1}(1-\gamma)}\right], \tag{C1}$$

which is a variant of equation (13). Rearranging terms in equation (C18) gives

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

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

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

Taking logs of both sides of equation (C3) gives

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

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

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

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

Columns (3)-(4) of Table 8 parameterize κ using two micro-foundations described in Section 6.2. These parameterizations complement our baseline parameterizations based on the forced sale discount, shown in columns (1)-(2). In both micro-foundations, we depart from our baseline framework by supposing that sellers experiencing a transaction failure receive an offer in month $t+1$ with probability $\phi < 1$.

In the first micro-foundation, we 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).

In the second micro-foundation, we allow the probability of attracting an offer in month $t+1$ to depend on κ according to the function $\phi(\kappa)$. This function reflects the Guren (2018) insight that a price cut increases the probability of sale, but larger price cuts only raise this probability incrementally: $\phi'(\kappa) > 0$, $\phi''(\kappa) < 0$. Therefore, in the context of the setup from Section 6.1, sellers choose κ to maximize sale probability,

$$\kappa = \phi'^{-1}(0). \tag{C10}$$

Appendix C.4 describes how we parameterize κ using equation (C10).

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,

$$\frac{\partial \mu}{\partial q} = \frac{e^{\kappa(\gamma-1)} - 1}{\gamma - 1}, \quad \frac{\partial \mu}{\partial \kappa} = q e^{\kappa(\gamma-1)}, \quad \frac{\partial \mu}{\partial r} = e^{r(1-\gamma)}, \tag{C11}$$

all of which are positive. In addition,

$$\begin{aligned} \frac{\partial \mu}{\partial \gamma} &= \frac{(\gamma-1)\left[q\kappa e^{\kappa(\gamma-1)} + r e^{r(1-\gamma)}\right] - \left[q\left(e^{\kappa(\gamma-1)} - 1\right) - e^{r(1-\gamma)} + 1\right]}{(\gamma-1)^2} \\ &= \frac{q e^{\kappa(\gamma-1)}\left[\kappa(\gamma-1) - 1\right] + e^{r(1-\gamma)}\left[1 + (\gamma-1) r\right] - (1-q)}{(\gamma-1)^2}. \end{aligned} \tag{C12}$$

A sufficient condition for equation (C12) to be positive is

$$q e^{\kappa(\gamma-1)}\left[\kappa(\gamma-1) - 1\right] > 1 - q, \tag{C13}$$

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,

$$\frac{\partial \mu}{\partial \xi} = 1, \quad \frac{\partial \mu}{\partial \ell} = \frac{\partial \mu}{\partial \delta} = \frac{q\left(1 - e^{-\kappa}\right)}{\left[e^{-\kappa} - \ell - \delta\right]^2}\left(\frac{1 - \ell - \delta}{e^{-\kappa} - \ell - \delta}\right)^{\gamma-2}, \tag{C14}$$

all of which are positive.

C.4 Calibration Details

As stated in Section 6.3, the externally calibrated parameters are q, κ, ξ, ℓ, and δ. 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^{-a}), and the simultaneous down payment (δe^{-a}). 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

As mentioned Section 6.3, we parameterize $q = 6.3\%$ based on the share of transactions that were terminated divided by the share of transactions with a mortgaged buyer over 2015-2020 (NAR 2018, 2020). This parameterization may overstate the true probability of transaction failure, since cash-financed purchases can also fail. However, the most common reasons reported for delayed or terminated settlements are "issues related to obtaining financing" (37%), "appraisal issues" (18%), and "home inspection" (16%), all of which pertain more commonly to mortgage-financed purchases (NAR 2018, 2020).

The alternative parameterization of $q = 12\%$ is based on the average mortgage application denial rate over 1994-2016 according to HMDA data. This parameterization likely constitutes an upper bound, since 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 will, nevertheless, be associated with a mortgage denial rate greater than zero. 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 κ according to the two micro-foundations described in Appendix C.2.

First, based on the micro-foundation summarized in equation (C9), we parameterize $\kappa = 6.0\%$. This value comes from 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.[19] In addition, we parameterize $\phi = 55\%$ based on Gilbukh and Goldsmith-Pinkham (2019).

Second, based on the micro-foundation summarized in equation (C10), we parameterize $\kappa = 11.1\%$. Following Guren (2018), we specify the demand curve, ϕ, as a quadratic function. Then, using the coefficients of this quadratic function as estimated in Table 2 of Guren (2018), we solve for the price cut that maximizes the probability of sale. Accordingly, we parameterize $\kappa = 11.1\%$.

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

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 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 $\ell e^{-a} = 33.1\%$ 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 $\delta e^{-a} = 17.5\%$ 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 $a = 6\%$.

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 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. 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 Theoretical Premium with Non-Standard Preferences

The two non-standard preferences described in Section 7.6 are present focus (i.e., present bias) and reference dependence.

Present Focus

Suppose sellers exhibit present bias, a particular form of present focus. A seller in month t discounts utility in month $t + \Delta t$ by the factor

$$\tilde{\beta} e^{-\rho \Delta t}, \tag{C15}$$

where $\tilde{\beta}$ is the quasi-hyperbolic discount factor. As in Appendix C.1, define the consumption-equivalent quasi-hyperbolic discount factor β as the solution to

$$\tilde{\beta} e^{-\rho \Delta t} u(C) = u\left(\beta e^{-r\Delta t} C\right), \tag{C16}$$

where r is the consumption-equivalent discount rate defined in equation (C6). In particular, the composite discount factor is

$$\tilde{\beta} e^{-\rho \Delta t} = \left(\beta e^{-r\Delta t}\right)^{1-\gamma}. \tag{C17}$$

Therefore, under the simple model described in the text in which $q = \ell = \delta = \xi = a = 0$, the mortgage-cash premium solves

$$\frac{\left(P_{i,t}^{C}\right)^{1-\gamma}}{1-\gamma} = \tilde{\beta} e^{-\rho \Delta t} \frac{\left(P_{i,t}^{M}\right)^{1-\gamma}}{(1-\gamma)}, \tag{C18}$$

which, by a similar logic as in Appendix C.1, yields the following expression,

$$\mu = r\Delta t - \log(\beta), \tag{C19}$$

as shown in equation (18) of the text.

Reference Dependence

Suppose sellers form a reference point around the price $\overline{P}_{i,t} \equiv P_{i,t}^{C} e^{-\Lambda}$. They have gain-loss utility that depends on whether utility under the actual sale price, $P_{i,t}$, exceeds that under the reference point. Following the setup and notation in Kőszegi and Rabin (2006), this gain-loss utility enters additively into the standard flow utility functions shown in equations (10), (11), and (12). If $u(P_{i,t}) \geq u\left(\overline{P}_{i,t}\right)$ such that the seller experiences a gain, then her utility equals

$$u_{i,t}^{+} = \frac{(P_{i,t})^{1-\gamma}}{1-\gamma} + \eta \left[\frac{(P_{i,t})^{1-\gamma}}{1-\gamma} - \frac{\left(\overline{P}_{i,t}\right)^{1-\gamma}}{1-\gamma}\right], \tag{C20}$$

where $\eta \geq 0$ governs the relative importance of gain-loss flow utility, as opposed standard flow utility. If, instead, $u(P_{i,t}) < u\left(\overline{P}_{i,t}\right)$ such that the seller experiences a loss, then her utility equals

$$u_{i,t}^{-} = \frac{(P_{i,t})^{1-\gamma}}{1-\gamma} - \lambda\eta\left[\frac{\left(\overline{P}_{i,t}\right)^{1-\gamma}}{1-\gamma} - \frac{(P_{i,t})^{1-\gamma}}{1-\gamma}\right], \tag{C21}$$

where $\lambda \geq 1$ governs the degree of aversion to loss relative to the reference point. We introduce the following variable

$$\tilde{\lambda}^F = \begin{cases} 1, & P_{i,t}^F \geq \overline{P}_{i,t} \\ \lambda, & P_{i,t}^F < \overline{P}_{i,t} \end{cases}, \tag{C22}$$

for $F \in \{C, M, R\}$, which governs whether the price under cash financing (C), mortgage financing (M), or after a relisting (R) represents a gain ($\tilde{\lambda}^F = 1$) or a loss ($\tilde{\lambda}^F = \lambda$).

Under the simple model described in the text in which $r = \ell = \delta = \xi = a = 0$, the mortgage-cash premium solves

$$\begin{aligned}\frac{(1+\eta\tilde{\lambda}^C)\left(P_{i,t}^C\right)^{1-\gamma} - \eta\tilde{\lambda}^C\left(\overline{P}_{i,t}\right)^{1-\gamma}}{1-\gamma} &= \frac{(1+\eta\tilde{\lambda}^M)\left(P_{i,t}^M\right)^{1-\gamma} - \eta\tilde{\lambda}^M\left(\overline{P}_{i,t}\right)^{1-\gamma}}{(1-q)^{-1}(1-\gamma)} + \ldots \\ &\ldots + \frac{(1+\eta\tilde{\lambda}^R)\left(P_{i,t}^R\right)^{1-\gamma} - \eta\tilde{\lambda}^R\left(\overline{P}_{i,t}\right)^{1-\gamma}}{q^{-1}(1-\gamma)}.\end{aligned} \tag{C23}$$

Dividing by $\left(P_{i,t}^C\right)^{1-\gamma}(1-\gamma)^{-1}$ and rearranging terms gives

$$e^{\mu(1-\gamma)} = \frac{1+\eta\tilde{\lambda}^C + \eta e^{\Lambda(\gamma-1)}\left[q\tilde{\lambda}^R + (1-q)\tilde{\lambda}^M - \tilde{\lambda}^C\right] - q\left(1+\eta\tilde{\lambda}^R\right)e^{\kappa(\gamma-1)}}{(1-q)\left(1+\eta\tilde{\lambda}^M\right)}. \tag{C24}$$

Since $P_{i,t}^M \geq P_{i,t}^C \geq P_{i,t}^R$, there are three cases. First, suppose $\tilde{\lambda}^R = \tilde{\lambda}^C = \tilde{\lambda}^M$. In this case, all methods of financing result in a certain gain, or they all result in a certain loss. Dividing equation (C23) by $1+\eta\tilde{\lambda}^F$ and following a similar logic as in Appendix C.1, the mortgage-cash premium then equals

$$\mu = \frac{1}{\gamma-1}\left[q\left(e^{\kappa(\gamma-1)} - 1\right)\right]. \tag{C25}$$

Second, suppose $\tilde{\lambda}^R = \tilde{\lambda}^C = \lambda > 1 = \tilde{\lambda}^M$. In this case, a failed mortgaged transaction and an all-cash transaction both result in a loss, whereas a successful mortgaged transaction results in a gain. Equation (C23) then implies the following expression for the mortgage-cash premium

$$\begin{aligned}\mu = &\frac{1}{\gamma-1}\left[q\left(\frac{1+\eta\lambda}{1+\eta}e^{\kappa(\gamma-1)} - \frac{\eta(\lambda-1)}{1+\eta}e^{\Lambda(\gamma-1)} - 1\right)\right] - \ldots \\ &\ldots - \frac{1}{\gamma-1}\left[\left(\frac{1+\eta\lambda}{1+\eta} - \frac{\eta(\lambda-1)}{1+\eta}e^{\Lambda(\gamma-1)} - 1\right)\right],\end{aligned} \tag{C26}$$

or, more compactly,

$$\mu = \frac{1}{\gamma - 1}\left[q\left(e^{\kappa(\gamma-1)} - 1\right)\right] + ... \tag{C27}$$
$$... + q\frac{\eta}{1+\eta}(\lambda - 1)\left(\frac{e^{-\Lambda(1-\gamma)}}{1-\gamma} - \frac{e^{-\kappa(1-\gamma)}}{1-\gamma}\right) + \frac{\eta}{1+\eta}(\lambda - 1)\left(\frac{e^{-\Lambda(1-\gamma)}}{1-\gamma} - \frac{1}{1-\gamma}\right).$$

Notice that equation (C27) is increasing in λ if and only if

$$q\left(\frac{e^{-\Lambda(1-\gamma)}}{1-\gamma} - \frac{e^{-\kappa(1-\gamma)}}{1-\gamma}\right) \geq \left(\frac{e^{-\Lambda(1-\gamma)}}{1-\gamma} - \frac{1}{1-\gamma}\right), \tag{C28}$$

that is, if the expected utility loss relative to the reference point under mortgage financing exceeds the certain utility loss relative to the reference point under cash financing.

Third, suppose $\tilde{\lambda}^R = \lambda > 1 = \tilde{\lambda}^C = \tilde{\lambda}^M$. In this case, only a failed mortgaged transaction results in a loss. Again using equation (C23) and following a similar logic used to derive equation (C27), the mortgage-cash premium then equals

$$\mu = \frac{1}{\gamma - 1}\left[q\left(e^{\kappa(\gamma-1)} - 1\right)\right] + q\frac{\eta}{1+\eta}(\lambda - 1)\left(\frac{e^{-\Lambda(1-\gamma)}}{1-\gamma} - \frac{e^{-\kappa(1-\gamma)}}{1-\gamma}\right). \tag{C29}$$

We can now formalize the claim from the text that the mortgage-cash premium is generically increasing in the degree of loss aversion if and only if $\kappa > \Lambda > 0$.

Proposition C.1 *The mortgage-cash premium μ is generically increasing in λ if and only if $\kappa > \Lambda > 0$. In that case,*

$$\mu = \frac{1}{\gamma - 1}\left[q\left(e^{\kappa(\gamma-1)} - 1\right)\right] + q\frac{\eta}{1+\eta}(\lambda - 1)\left(\frac{e^{-\Lambda(1-\gamma)}}{1-\gamma} - \frac{e^{-\kappa(1-\gamma)}}{1-\gamma}\right).$$

Proof If $\kappa > \Lambda > 0$, then $\tilde{\lambda}^R = \lambda > 1 = \tilde{\lambda}^C = \tilde{\lambda}^M$ and the mortgage-cash premium equals the expression in equation (C29). According to equation (C29), μ is increasing in λ. In the opposite direction, the set of possible expressions for μ are equations (C25), (C27), and (C29). The only one of these equations in which μ is strictly increasing in λ for any parameterization is equation (C29), corresponding to the case $\tilde{\lambda}^R = \lambda > 1 = \tilde{\lambda}^C = \tilde{\lambda}^M$. This completes the proof.

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_1RJKGXQH4I8pDfg.

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 6.41 minutes to do so. We administered the survey in April 2021 and received data from 980 respondents, all of which come from U.S. homeowners. This period marks an unprecedented "seller's market", in that it features low interest rates, strong demand, and limited supply (Gascon and Haas 2020). Consequently, the mortgage-cash premium elicited through our survey likely provides a conservative estimate of the steady state premium. We retain respondents who spend between three and thirty minutes on the survey and are at least twenty years old, which reduces the sample size to 938, 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%; 36.3%; 41.3%; 18.0% and 3.2%. 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. First, 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.[20] Note that the quantity \$[$0.3 \times B$] corresponds to the variable $L_{i,t}$ in our framework, such that $\ell e^{-a} = 0.3$, 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 5 or 7, 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 to straddle the average closing period of 47 days for such transactions (Ellie Mae 2012). Note that the quantity \$[$0.15 \times B$] maps to $D_{i,t}$, such that $\delta e^{-a} = 0.15$ to target the value from Table 7.

[20]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."* Also, the share of the list price at which the home was bought ten years ago randomly equals either 0.5 or 0.75, with 50% probability for each value. We find no significant difference in results between respondents assigned to these two purchase price factors.

The subsequent three screens describe the set of potential homebuyers. 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 next 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, it will take two weeks to close the transaction. 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."* While a two week closing period for cash-financed purchases departs from our stark theoretical framework, it encourages experienced home sellers to take the survey seriously because it matches anecdotal evidence on the actual cash-financed closing period (e.g., Zillow 2021).

On the last screen describing the thought experiment, we describe 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."* As just mentioned, the parameter T randomly equals either 5 or 7. This randomization allows us to assess the scope for non-exponential discounting, as we do in Table 11.

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]. So, both buyers offer to pay your list price. Which offer would you accept today?"*

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 two most important reasons."*. Respondents can select two options from the following set:

1. *"If the Mortgaged Buyer backs out and I relist at a lower price, I may not be able to cover my expenses."*
2. *"If the Mortgaged Buyer backs out and I relist at a lower price, I'd regret that I could have received a higher price from the Cash Buyer."*
3. *"If the Mortgaged Buyer backs out and I relist at a lower price, I might not realize an attractive financial return on my home."*
4. *"Even if I was completely certain that the Mortgaged Buyer would follow through, they would still take a longer time to close."*

5. *"Other (please describe)"*

Importantly, the order of the first four of these options is randomized, and so the results do not confound tendencies to select options that appear at the top of a list. The first option aims to recover the joint role of risk aversion (i.e., γ) and the need to meet fixed obligations (i.e., ℓ, δ). The second option aims to recover the role of reference dependence around the non-accepted offer. The third option aims to recover the role of realized financial return. The fourth option aims to recover the role of impatience (i.e., r), as distinct from risk aversion. Finally, 8% of respondents describe an "other" answer. As suggested by Appendix Figure A4, however, most of these responses tend to corroborate one of the four options listed in the question.

The following screen contains a series of questions that elicit a respondent's mortgage-cash premium. 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 5% to 20% in increments of 5 pps. We emphasize that respondents are shown the price differentials in both levels and percentages to account for non-proportional thinking.

The majority of respondents switch from preferring the all-cash offer to the mortgaged offer once $\tilde{\mu}$ passes some 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,

$$\begin{aligned} \mathit{Premium}_k = \frac{1}{2} \big[&\min \{ \tilde{\mu} | \, \mathit{Mortgaged\ Offer} \succ \mathit{All\text{-}Cash\ Offer} \} + ... \\ &... + \max \{ \tilde{\mu} | \, \mathit{All\text{-}Cash\ Offer} \succ \mathit{Mortgaged\ Offer} \} \big] . \end{aligned} \tag{D1}$$

We assign a mortgage-cash premium of 22.5% for respondents who still prefer the all-cash offer at a price spread of 20%, such that the mortgage-cash premium either equals zero or ranges from 2.5% to 22.5% in increments of 5 pps.

The final three screens in the survey's core elicit the respondent's beliefs about transaction risk. 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 20% on a sliding scale. Similarly, we then ask respondents: *"Homes with lower listing prices typically sell more quickly. At what price would you relist your home in [T] weeks 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. The difference between the log of the answer and the log of the list price determines the value of κ shown in Table 10. Finally, we assess the role of non-proportional thinking by asking respondents the percent price cut corresponding to their answer from the previous question: *"If you indeed need to relist your home at \$[$e^{-\kappa} \times B$] because the transaction fails, by what percent would you need to cut the list price relative to your current list price?"* The variable *High Numeracy*$_k$ indicates if the answer lies within 5 pps of the value of κ implied by the dollar value by which they would reduce the price. This condition applies to 77% of respondents.

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:

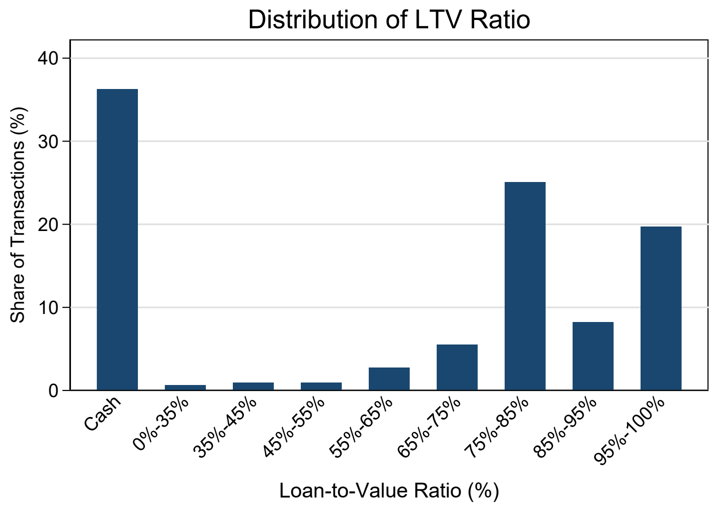
1. *"I am always willing to take risks."*
2. *"I am usually willing to take risks, but I am sometimes reluctant to do so."*
3. *"I am usually reluctant to take risks, but I am sometimes willing to do so."*
4. *"I am never willing to take risks."*

The variable *High Risk Aversion*$_k$ indicates if respondent k chooses either the third or the fourth option from this list.

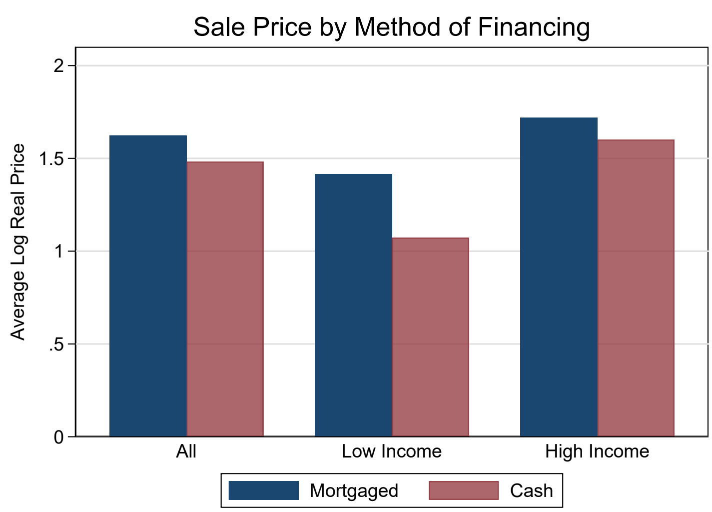
The remaining six questions assess the respondent's experience in real estate markets. In particular, we ask respondents: whether they own or rent the home they currently live in; whether they own any investment properties; whether they are a professional real estate agent; whether they have ever sold a home and, if so, the number of occasions; whether they have ever sold a home to a mortgaged buyer and, if so, the number of occasions; and, for respondents who have previously sold a home, the similarity between the survey's thought experiment and their own sale experience.

Additional Figures and Tables

Figure A1: Robustness of Facts to Alternative Dataset



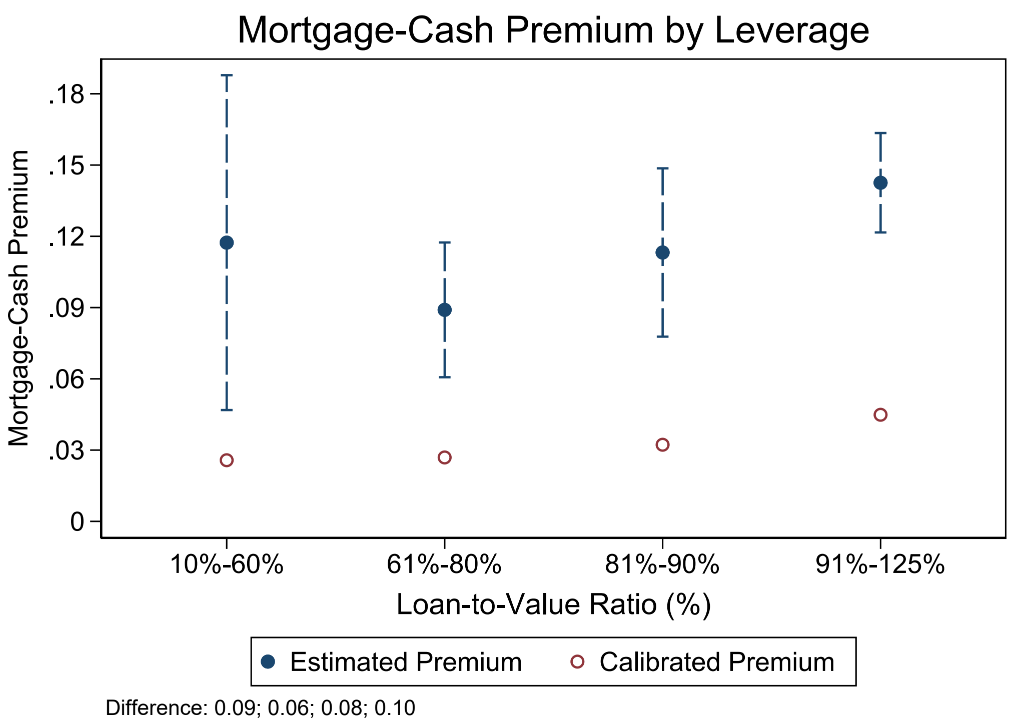
(a) Distribution of Loan-to-Value Ratio



(b) Sale Price by Method of Financing

Note: This figure produces plots analogous to those in Figure 1 using the offer-level dataset, which assesses the robustness of the facts in Section 3. Panel (a) plots the distribution of loan-to-value ratios across home purchases (i.e., accepted offers) over the period from May 2020 through October 2020. Panel (b) plots the average log sales price in hundreds of thousands of 2020 dollars for purchases financed by a mortgage (Mortgaged) and exclusively with cash (Cash). Panel (b) excludes purchases with less than three offers and adjusts variables for seasonality. The remaining notes are the same as in Figure 1.

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,

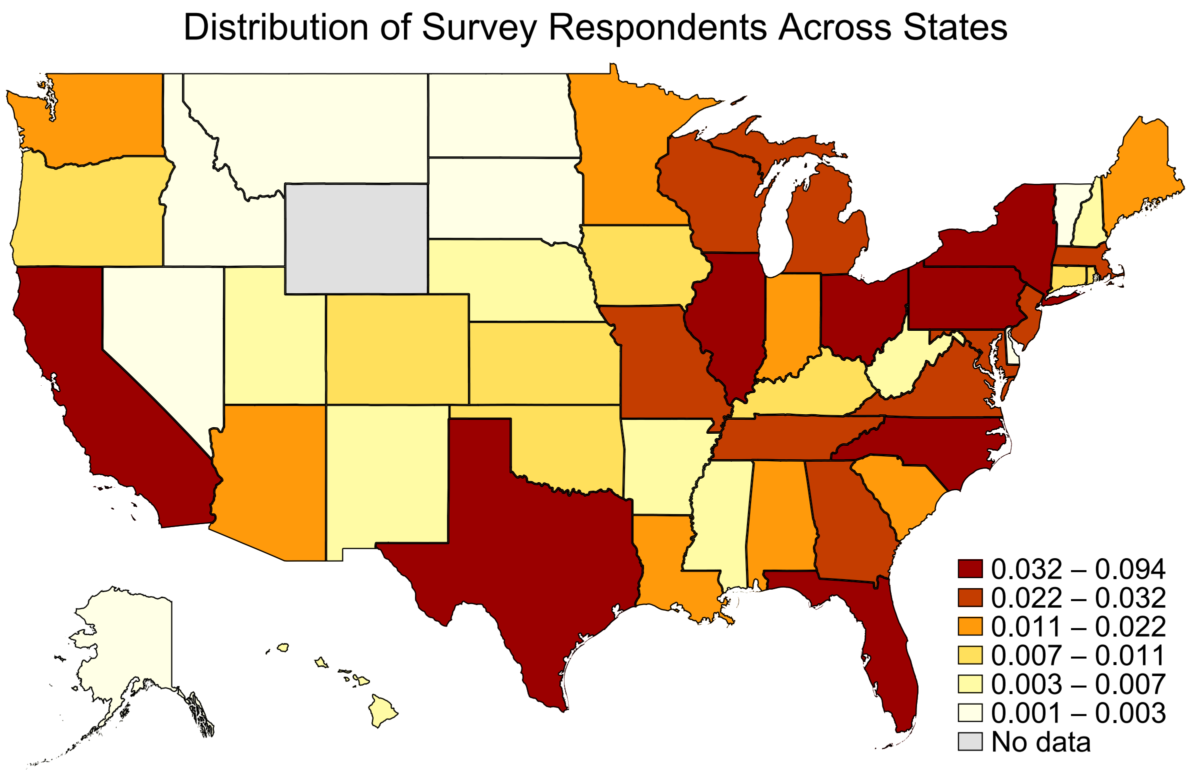
$$\log\left(Price_{i,t}\right) = \mu Mortgaged_{i,t} + \psi X_{i,t} + \zeta_{z(i),t} + \alpha_i + \epsilon_{i,t}, \quad (i,t) \in \mathcal{L}$$

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.

Figure A4: Motivation for Preferring All-Cash Offers from Survey Free Response

Note: This figure plots a word cloud for survey respondents' free response to the question: *"Why would you prefer to sell your home to the Cash Buyer rather than the Mortgaged Buyer?"* The size of each word corresponds to the frequency with which it is used in respondents' answers. The set of responses have been preprocessed to retain nouns, verbs, adverbs, and adjectives and to harmonize variant spellings, punctuation, and synonymous words. Data are from the survey described in Section 7.

Table A1: Observed Transaction Characteristics by the Method of Financing

	Mean		
	Cash	Mortgaged	Difference
	(1)	(2)	(3)
$Age_{i,t}$	28.906	28.705	-0.201
			(0.000)
$Months\ Owned_{i,t}$	27.617	45.449	17.832
			(0.000)
$Non\ Institutional_{i,t}$	0.819	0.888	0.069
			(0.000)
$Non\ Foreign_{i,t}$	0.825	0.837	0.012
			(0.000)
Number of Observations	426,256	426,256	426,256

Note: P-values are in parentheses. This table summarizes observed characteristics of cash and mortgage-financed purchaes that we use to partition the sample in Table 2. Subscripts i and t index property and month. The variable $Age_{i,t}$ is defined in the note to Table 1; $Months\ Owned_{i,t}$ is the number of months from month t until property i is sold again; $Non\ Institutional_{i,t}$ indicates whether neither the buyer nor the seller is an institution; and $Non\ Foreign_{i,t}$ indicates whether neither the buyer nor the seller has a foreign address. Columns (1)-(2) summarize the average value of each variable for cash and mortgage-financed purchases, respectively. Column (3) tests for a difference in mean. The remaining notes are the same as in Table 2.

Table A2: Robustness to Properties Without a Repeat Sale

Outcome:	$\log(Price_{i,t})$		
	(1)	(2)	(3)
$Mortgaged_{i,t}$	0.161	0.176	0.099
	(0.000)	(0.000)	(0.000)
Sample	Full	No Repeat Sale	Repeat Sale
Zip Code-Month FE	Yes	Yes	Yes
Hedonic-Month FE	Yes	Yes	Yes
Property FE	No	No	No
R-squared	0.582	0.573	0.714
Number of Observations	2,254,389	1,821,253	426,256

Note: P-values are in parentheses. This table estimates equation (4) after including properties without a repeat sale in the sample, which assesses the external validity of the baseline results. Subscripts i and t index property and month. The samples in columns (1)-(3) include all properties regardless of whether they experience a repeat sale, properties with no repeat sale, and properties with a repeat sale, respectively. The remaining notes are the same as in Table 2.

Table A3: Robustness to Semi-Structural Estimator from Bajari et al. (2012)

Outcome:	$\log(Price_{i,t})$	
	(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,

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

where the first-stage regression equation is

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

and where n denotes holding period; and $\varrho_t(n), \varphi_t^M(n)$, and $\varphi_t^P(n)$ 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 A4: 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.